

Probabilités

Céline Esser

Université de Liège, Année académique 2025 – 2026

Table des matières

1	Variables aléatoires	1
1.1	Définition	1
1.2	Loi d'une variable aléatoire	3
1.3	Fonction de répartition	6
1.4	Lois usuelles	12
1.4.1	Distributions univariées discrètes	12
1.4.2	Distributions univariées continues	17
1.5	Vecteurs aléatoires	22
1.6	Lois multivariées usuelles	28
1.6.1	Loi multinomiale	29
1.6.2	Loi normale multivariée	29
2	Espérance mathématique	31
2.1	Espérance d'une variable aléatoire	31
2.2	Moments et variance	36
2.3	Covariance et corrélation	40
2.4	Moments de vecteurs aléatoires	44
2.5	Formules et inégalités liées à l'espérance	45
2.6	Fonction caractéristique	48
3	Indépendance	52
3.1	Indépendance d'événements	52
3.2	Indépendance de σ -algèbres	54
3.3	Indépendance de variables aléatoires	57
3.4	Loi du 0-1 de Kolmogorov	61
3.5	Indépendance et espérance	62
3.6	Sommes de variables aléatoires indépendantes	64
3.7	Vecteurs aléatoires gaussiens	68

4	Espérances conditionnelles et martingales	72
4.1	Introduction et motivation	72
4.1.1	Lois conditionnelles discrètes	72
4.1.2	Lois conditionnelles continues	76
4.1.3	Autres exemples	78
4.1.4	Espérance et projection	80
4.2	L'espérance conditionnelle	80
4.2.1	σ -algèbre et information	80
4.2.2	Espérance conditionnelle à une σ -algèbre	82
4.3	Propriétés de l'espérance conditionnelle	84
4.4	Calcul d'espérances conditionnelles	90
4.4.1	Accord avec les définitions intuitives	90
4.4.2	Cas d'une σ -algèbre discrète	93
4.4.3	Conditionnement gaussien	94
4.5	Martingales	97
4.6	Temps d'arrêt et martingales arrêtées	103
5	Convergences	107
5.1	Convergence presque sûre	107
5.2	Convergence en moyenne et en norme L^p	110
5.3	Convergence en probabilité	112
5.4	Lois des grands nombres	115
5.5	Convergence en loi et théorème central limite	126
5.6	Extension au cas multivarié et théorèmes de Slutsky	139
5.7	Convergence dans le cas gaussien	142
A	Introduction et rappels de théorie de la mesure	144
A.1	Approche intuitive de la notion de probabilité	144
A.2	σ -algèbres et applications mesurables	147
A.3	Mesures	149
A.4	Classes de Dynkin et Lemme de la classe monotone	150
A.5	Intégration et théorèmes limites	152
A.6	Mesure image	155
A.7	Le théorème de Radon-Nikodym	156
A.8	Mesure produit et théorème de Tonelli-Fubini	158
A.9	Espaces L^p	160

B	Notions fondamentales de la théorie des probabilités	163
B.1	Espaces probabilisés	163
B.2	Mesure de probabilité conditionnelle	165
B.3	Indépendance d'événements	169
C	Théorème de Radon-Nikodym	172
C.1	Densités	172
C.2	Mesures absolument continues et singulières	173
C.3	Décomposition de Lebesgue et Théorème de Radon-Nikodym	174
	Références	184

Avant propos

Ces notes de cours ont été principalement rédigées durant l'année académique 2022-2023 et sont destinées aux étudiants du cours de Probabilités dispensé dans le bloc 3 du cursus en Sciences Mathématiques.

L'objectif de ce cours est de fournir aux étudiants une compréhension approfondie des concepts fondamentaux de la théorie des probabilités. Cette solide base en théorie des probabilités constituera une compétence précieuse, préparant les étudiants à aborder des domaines tels que la statistique et l'analyse stochastique.

Le premier chapitre est destiné à introduire la notion de variable aléatoire et de loi. Nous montrons que la loi d'une variable aléatoire est une mesure de probabilité sur \mathbb{R} et que réciproquement, toute mesure de probabilité sur \mathbb{R} définit une variable aléatoire. Cela permet d'étudier les variables aléatoires via les mesures de probabilités. Nous montrons également que les lois sont entièrement caractérisées par leur fonction de répartition, ce qui fournit une troisième manière équivalente de travailler avec un même objet. Nous passons également en revue les lois classiques.

Le chapitre 2 est centré autour de la notion d'espérance. L'espérance d'une variable aléatoire n'est rien d'autre que l'intégrale de cette fonction dans la mesure de probabilité de l'espace sur lequel elle est définie. Bien que la notion d'intégrale ait été intensivement étudiée dans le cours de Calcul Intégral, cette notion étant au cœur de la théorie des probabilités, nous y consacrons un chapitre dans lequel nous introduisons le vocabulaire des probabilités et traduisons les résultats de la théorie de la mesure dans ce langage.

Dans le chapitre 3, nous revenons sur la notion d'indépendance, en étendant la définition d'indépendance d'événements au cas des σ -algèbres. Cela permet d'étudier l'indépendance de variables aléatoires et de démontrer la loi du 0-1 de Kolmogorov. Nous montrons ensuite comment l'espérance permet de caractériser l'indépendance.

L'objectif du chapitre 4 est d'étudier l'espérance conditionnelle. La définition de celle-ci étant peu intuitive, nous commençons par une approche intuitive sur des exemples discrets ou continus. Après avoir défini proprement l'espérance conditionnelle, nous étudions ensuite les propriétés qui découlent immédiatement de sa définition. La dernière section propose une très brève introduction à la théorie des martingales.

Le chapitre 5 est dédié à l'étude de la convergence d'une suite de variables aléatoires. Nous commençons par étudier la convergence presque sûre et ses liens avec la convergence en norme L^p et en probabilité. Nous démontrons ensuite la loi des grands nombres et l'une de ses applications importantes qui permet de justifier l'approche fréquentiste des probabilités. Dans un second temps, nous introduisons la convergence en loi et ses caractérisations équivalentes. Cela nous permet de démontrer l'un des théorèmes les plus importants de la théorie des probabilités, le théorème central limite.

Dans les chapitres repris en annexe, nous rappelons l'approche naïve de la notion de probabilité et nous motivons l'intérêt de la théorie de la mesure dans l'étude des probabilités. Ce chapitre contient de nombreux rappels du cours de Calcul Intégral qui seront utilisés de manière récurrente dans le cours. La deuxième annexe rappelle une étude élémentaire des probabilités, en rappelant le vocabulaire approprié introduit dans le cours de Statistique Descriptive. Par rapport au cours de Calcul Intégral, nous rappelons la notion d'indépendance d'événements, qui est indispensable et centrale en théorie des probabilités. Enfin, dans la troisième annexe, nous proposons une preuve du Théorème de Radon-Nikodym, non-vue dans le cours de Calcul Intégral mais indispensable dans le cadre de la théorie des probabilités.

Ces notes de cours ont été rédigées sur la base de nombreux échanges avec Yvik Swan ainsi que de son cours donné à l'ULB. Je l'en remercie chaleureusement. Je remercie également les étudiants du bloc 3 en Sciences Mathématiques de l'année académique 2022-2023 qui m'ont signalé de nombreuses coquilles présentes dans la première version des notes.

Chapitre 1

Variables aléatoires

Les variables aléatoires permettent de modéliser de manière rigoureuse et systématique des phénomènes aléatoires en associant à chaque issue possible d'une expérience un nombre réel. Elles offrent un cadre mathématique unifié pour l'étude d'expériences variées, facilitant le calcul des probabilités associées à divers événements d'intérêt.

1.1 Définition

Considérons une expérience aléatoire et l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ associé. L'outil de base que nous allons utiliser et étudier est la *variable aléatoire* qui associe aux issues ω de l'expérience (qui, pour rappel, peuvent être de *toute* nature) des nombres réels.

Définition 1.1.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une application mesurable

$$X : \Omega \rightarrow \mathbb{R}$$

est appelée une *variable aléatoire (réelle)*.

Ainsi, une variable aléatoire est une application $X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega)$ telle que, pour tout Borélien $B \in \mathcal{B}$, on a

$$\{X \in B\} \in \mathcal{F} \tag{1.1}$$

où la notation $\{X \in B\}$ est une *convention* usuelle pour représenter l'ensemble $X^{-1}(B)$, i.e.

$$\{X \in B\} = X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

L'événement $\{X \in B\}$ aura donc une probabilité. On omet généralement les accolades dans l'écriture de sa probabilité :

$$\mathbb{P}(X \in B) = \mathbb{P}(\{X \in B\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

Il est d'usage de représenter les variables aléatoires par les lettres X, Y, Z, \dots

Remarque 1.1.2 Si $\mathcal{F} = \mathcal{P}(\Omega)$ alors toute fonction réelle sur Ω est une variable aléatoire.

Remarque 1.1.3 Ce n'est pas la fonction qui est aléatoire : quand ω est observé on *sait* quelle valeur prendra X . Toutefois, ω étant lui même choisi de façon aléatoire (avec probabilité dictée par \mathbb{P}), on ne sait pas quelle valeur prendra X ; la valeur de X est donc malgré tout aléatoire.

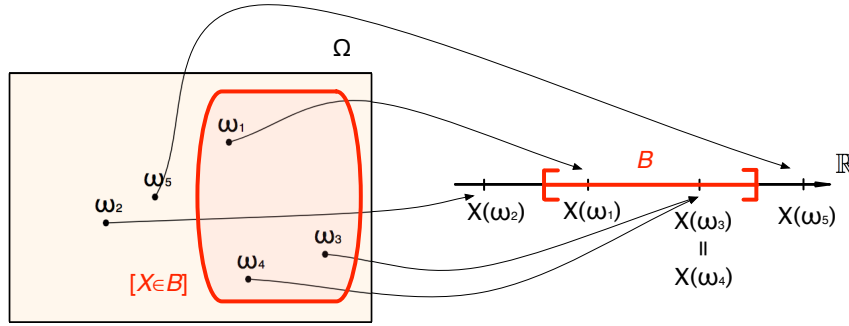


FIGURE 1.1 – Une variable aléatoire transforme des ω en nombres réels ; tout intervalle provient d’un événement de \mathcal{F} .

Exemple 1.1.4 Considérons l’expérience du lancer de dé de l’exemple B.1.3. Pour étudier la parité du résultat, on considère la fonction donnée par

$$X : \Omega \rightarrow \mathbb{R} : i \mapsto X(i) = i - 2 \left\lfloor \frac{i}{2} \right\rfloor.$$

Puisque $\mathcal{F} = \mathcal{P}(\Omega)$, la condition (1.1) est satisfaite et X est une variable aléatoire. L’événement $\{X = 0\}$ correspond à l’ensemble $A = \{2, 4, 6\}$.

Exemple 1.1.5 Considérons l’expérience du lancé de deux dés distinguables dont l’espace associé est $(\Omega, \mathcal{F}) = (\{(i, j) : i, j \in \{1, \dots, 6\}\}, \mathcal{P}(\Omega))$. Soit X la somme des résultats de chaque dé, i.e.

$$X : \Omega \rightarrow \mathbb{R} : (i, j) \mapsto X(i, j) = i + j.$$

Puisque $\mathcal{A} = \mathcal{P}(\Omega)$, la condition (1.1) est satisfaite et X est une variable aléatoire. L’événement $\{X \geq 11\}$ correspond à l’ensemble $A = \{(5, 6), (6, 5), (6, 6)\}$.

Exemple 1.1.6 Considérons l’expérience qui consiste à lancer une pièce de monnaie autant de fois que nécessaire pour obtenir pile. L’espace associé est $(\Omega, \mathcal{F}) = (\{p, (f, p), (f, f, p), \dots\}, \mathcal{P}(\Omega))$. Soit X le nombre de lancers avant observation d’un pile, i.e.

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) = \#f.$$

Puisque $\mathcal{F} = \mathcal{P}(\Omega)$, la condition (1.1) est satisfaite et X est une variable aléatoire.

Le résultat suivant est une conséquence directe de la Proposition A.2.10.

Théorème 1.1.7 Soit (Ω, \mathcal{F}) un espace mesurable. Alors $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire si et seulement si $X^{-1}([-\infty, x]) \in \mathcal{F}$ pour tout $x \in \mathbb{R}$.

Corollaire 1.1.8 Soit (Ω, \mathcal{F}) un espace mesurable et soit $X : \Omega \rightarrow \mathbb{R}$ une application telle que $X(\Omega)$ est dénombrable. Alors X est une variable aléatoire si et seulement si pour tout $x \in X(\Omega)$, on a $X^{-1}(\{x\}) \in \mathcal{F}$.

Démonstration : Il suffit de remarquer que $X^{-1}(]-\infty, x]) = \bigcup_{y \in X(\Omega), y \leq x} X^{-1}(\{y\})$. ■

Exemple 1.1.9 Considérons l'expérience qui consiste à mesurer le temps d'attente du 48 à un arrêt de bus. L'espace probabilisé associé est $(\Omega, \mathcal{F}) = (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$.

— Soit X l'indicateur de retard donné par

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) = \begin{cases} 1 & \text{si } \omega > 1 \\ 0 & \text{sinon.} \end{cases}$$

On remarque que

$$X^{-1}(\{1\}) =]1, +\infty[\text{ et } X^{-1}(\{0\}) =]0, 1[.$$

Les deux ensembles sont des boréliens, ce qui suffit pour que la condition (1.1) soit vérifiée. Ainsi, X est bien une variable aléatoire.

— Soit X le temps d'attente, i.e.

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) = \omega.$$

On a $X^{-1}(]-\infty, x]) =]-\infty, x] \in \mathcal{B}(\mathbb{R}^+)$ pour tout x . Ainsi, X est une variable aléatoire.

— Soit X le nombre de minutes d'attente ; $X(\Omega) = \mathbb{N}$. A nouveau on vérifie aisément qu'il s'agit d'une variable aléatoire puisque $X^{-1}(\{n\}) = [n, n + 1[$.

Remarque 1.1.10 Parfois, on contourne ces questions de mesurabilité en travaillant avec la σ -algèbre engendrée par X . En général, il est compliqué de construire $\sigma(X)$ explicitement. On se contentera donc de son existence, et on ne la spécifiera pas.

On pourrait également considérer des variables aléatoires à valeurs dans \mathbb{R}^d , ou plus généralement dans un espace E . La définition s'étend comme suit.

Définition 1.1.11 Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et (E, \mathcal{E}) un espace mesurable. Une application mesurable $X : \Omega \rightarrow E$ est appelée une *variable aléatoire* à valeurs dans E .

L'espace d'arrivée (E, \mathcal{E}) est l'espace d'états. Si $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ alors X est une *variable aléatoire réelle*, si $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ alors X est un *vecteur aléatoire*. On y reviendra à la fin du chapitre.

Remarque 1.1.12 Il est parfois utile ou nécessaire de considérer des variables aléatoires à valeurs dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Dans ce cas, on peut avoir $\mathbb{P}(X = +\infty) > 0$. En revanche on a $\mathbb{P}(-\infty \leq X \leq +\infty) = 1$.

1.2 Loi d'une variable aléatoire

Comme toute variable aléatoire est une fonction mesurable, on peut s'intéresser à la mesure qu'elle définit sur $(\mathbb{R}, \mathcal{B})$.

Définition 1.2.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit X une variable aléatoire. La *loi* de X est la mesure image \mathbb{P}_X de \mathbb{P} par X . C'est donc la mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$ définie par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B)$$

pour tout ensemble $B \in \mathcal{B}$. On l'appelle également la *distribution* de la variable aléatoire X .

On utilise la notation $X \sim \mu$ pour exprimer que $\mathbb{P}_X = \mu$.

Intuitivement, la loi de X est une description de toutes les probabilités des événements liés à X . En pratique, l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ est un peu “mythique”. Si on se donne une mesure de probabilité μ sur \mathbb{R} , on peut toujours l'écrire comme la loi image d'une application mesurable : il suffit de prendre $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$, $\mathbb{P} = \mu$ et de considérer la variable aléatoire $X = \text{id}$. Ainsi, toute mesure de probabilité sur \mathbb{R} est la loi d'une variable aléatoire. En général, on s'intéresse uniquement à la mesure image et on explicite rarement l'espace $(\Omega, \mathcal{F}, \mathbb{P})$. Par exemple, on écrit “soit X une variable aléatoire de loi μ ” pour dire “soit X une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que $\mathbb{P}_X = \mu$ ”.

Définition 1.2.2 Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ et Y une variable aléatoire sur $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. On dit que X et Y sont *de même loi* si $\mathbb{P}_X(B) = \mathbb{P}_Y(B)$ pour tout $B \in \mathcal{B}$. On dit également que X et Y sont *identiquement distribuées* et on note $X \stackrel{\mathcal{L}}{=} Y$.

Remarquons que l'énoncé de la définition 1.2.2 ne fait pas mention des espaces fondamentaux ou des σ -algèbres sous-jacentes et seules comptent les activités dans \mathbb{R} . En particulier, deux variables aléatoires peuvent être égales en loi *sans être définies sur le même espace* : X est défini implicitement sur “son” $(\Omega_X, \mathcal{F}_X)$ et Y sur “son” $(\Omega_Y, \mathcal{F}_Y)$, où on a donc que $\mathcal{F}_X \supset \sigma(X)$ et $\mathcal{F}_Y \supset \sigma(Y)$.

Exemple 1.2.3 Soit X la variable aléatoire qui rend 1 si on répond correctement à un QCM avec 3 distracteurs et 0 sinon. Alors, les valeurs prises par X sont données par l'ensemble $\{0, 1\}$ et on a $\mathbb{P}(X = 0) = \frac{3}{4} = 1 - \mathbb{P}(X = 1)$. Soit Y la variable aléatoire qui rend 1 si on tire un coeur d'un jeu de cartes classique et 0 sinon. Alors, les valeurs prises par Y sont données par $\{0, 1\}$ et on a $\tilde{\mathbb{P}}(Y = 0) = \frac{3}{4} = 1 - \tilde{\mathbb{P}}(Y = 1)$. Il est clair que $X \stackrel{\mathcal{L}}{=} Y$ sans pour autant qu'ils ne vivent sur le même espace. On peut aussi définir une troisième variable aléatoire de même loi vivant sur un espace probabilisé continu : on prend l'espace $([0, 1], \mathcal{B}([0, 1]), \lambda|_{[0, 1]})$ et la variable aléatoire $Z = \mathbb{1}_{[0, \frac{1}{4}]}$.

Remarquons également que, même si elles sont définies sur le même espace probabilisé, l'égalité en loi des variables aléatoires n'implique pas leur égalité \mathbb{P} -presque partout.

Exemple 1.2.4 Soit X le nombre de “pile” observés lors de 5 lancers consécutifs indépendants d'une pièce de monnaie. Soit Y le nombre de “face” observés lors des mêmes 5 lancers. Alors X et Y sont identiquement distribuées, alors que $\mathbb{P}(X = Y) = 0$.

Bien souvent, on met en évidence deux classes de variables aléatoires : les variables aléatoires dont la loi est discrète (combinaison linéaire finie ou dénombrable de masses de Dirac) et les variables aléatoires dont la loi est absolument continue par rapport à la mesure de Lebesgue. Comme signalé dans la décomposition de Lebesgue donnée dans le Théorème A.7.9, cette classification n'est pas dichotomique : il existe des lois qui ne sont ni discrètes, ni absolument continues par rapport à la mesure de Lebesgue (ni des combinaisons des deux).

Définition 1.2.5 (Lois discrètes) Une variable aléatoire X est *discrète* si sa loi \mathbb{P}_X est discrète, c'est-à-dire s'il existe une suite $(p_n)_{n \in \mathbb{N}}$ de réels positifs ou nuls et une suite $(x_n)_{n \in \mathbb{N}}$

de réels tels que

$$\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}.$$

Le résultat suivant est immédiat.

Proposition 1.2.6 Soit X une variable aléatoire discrète de loi $\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}$. Alors

- $0 \leq p_n \leq 1$ pour tout $n \in \mathbb{N}$,
- $\sum_{n \in \mathbb{N}} p_n = 1$,
- $p_n = \mathbb{P}(X = x_n)$ pour tout $n \in \mathbb{N}$.

On appelle la fonction

$$P_X : x \in \mathbb{R} \mapsto \mathbb{P}(X = x)$$

la fonction de masse de X .

Il suffit bien sûr de connaître la fonction de masse sur D pour connaître la loi de X . Plus précisément, si $\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}$, on calcule toutes les probabilités liées à X de la manière suivante : si $B \in \mathcal{B}$, on a

$$\mathbb{P}(X \in B) = \sum_{n: x_n \in B} p_n = \sum_{n: x_n \in B} \mathbb{P}(X = x_n).$$

Spécifier une loi discrète peut donc se faire en spécifiant la suite des réels $(x_n)_{n \in \mathbb{N}}$ ainsi que leurs poids $(p_n)_{n \in \mathbb{N}}$ associés. Par métonymie, la donnée des $(x_n, p_n)_{n \in \mathbb{N}}$ est également appelée la loi de X .

Proposition 1.2.7 Une variable aléatoire X est discrète si et seulement si il existe un ensemble dénombrable $D \subseteq \mathbb{R}$ tel que $\mathbb{P}_X(D) = 1$.

Démonstration : Si X est discrète de loi $\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}$, alors l'ensemble dénombrable $D = \{x_n : n \in \mathbb{N}\}$ est tel que $\mathbb{P}(X \in D) = \mathbb{P}_X(D) = 1$. Remarquons qu'on aurait pu prendre plus précisément $D = \{x_n : p_n > 0\}$. Pour la réciproque, comme D est dénombrable et $\mathbb{P}_X(D^c) = 0$, on a successivement

$$\mathbb{P}_X(A) = \mathbb{P}_X(A \cap D) + \mathbb{P}_X(A \cap D^c) = \sum_{x \in D} \mathbb{P}_X(A \cap \{x\}) = \sum_{x \in D} \mathbb{P}_X(\{x\}) \delta_x(A)$$

pour tout $A \in \mathcal{B}$. ■

Définition 1.2.8 (Lois continues) Une variable aléatoire X est *continue* si sa loi \mathbb{P}_X est absolument continue par rapport à la mesure de Lebesgue, c'est-à-dire s'il existe une application mesurable $f : \mathbb{R} \rightarrow [0, +\infty]$ telle que

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B f(x) dx$$

pour tout ensemble $B \in \mathcal{B}(\mathbb{R})$. Dans ce cas, la fonction f est unique à un ensemble λ -négligeable près et est appelée la *densité* de X .

En particulier, on a

$$\int_{\mathbb{R}} f(x) dx = 1$$

et donc f est intégrable. On peut donc supposer qu'elle est à valeurs dans $[0, +\infty[$ puisque cette propriété est vérifiée λ -pp. De plus, pour tous réels $a < b$,

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx.$$

On remarque également que $\mathbb{P}(X = a) = 0$ pour tout $a \in \mathbb{R}$.

Nous verrons dans la Section 1.6 des exemples classiques de lois discrètes ou continues.

1.3 Fonction de répartition

Nous avons vu que la loi d'une variable aléatoire discrète est entièrement déterminée par sa fonction de masse, tandis que la loi d'une variable aléatoire continue est entièrement déterminée par sa fonction de densité. En général, les variables aléatoires ne sont ni discrètes, ni continues. On ne peut donc pas spécifier leur loi via les fonctions de masse ou de densité. Dans cette section nous associons à toute variable aléatoire, quelle que soit sa nature, une nouvelle fonction qui caractérise sa loi.

Définition 1.3.1 Soit X une variable aléatoire. La fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x])$$

pour tout $x \in \mathbb{R}$ est appelée la *fonction de répartition* de X .

Puisque les demi-droites engendrent la σ -algèbre de Borel, le fait que la fonction de répartition d'une variable aléatoire caractérise entièrement sa loi est une conséquence directe du Lemme de la classe monotone rappelé dans la Proposition A.4.3.

Théorème 1.3.2 Deux variables aléatoires X et Y sont identiquement distribuées si et seulement si $F_X = F_Y$.

Par conséquent, on écrira souvent $X \sim F$ pour dire qu'on considère une variable aléatoire dont la fonction de répartition est F .

Exemple 1.3.3 Si X est une variable aléatoire qui prend presque sûrement la valeur a (i.e. $\mathbb{P}_X = \delta_a$), alors $F_X(x) = \mathbb{1}_{[a, +\infty[}$.

Exemple 1.3.4 Si X est une variable aléatoire discrète et si D est un ensemble dénombrable tel que $\mathbb{P}_X(D) = 1$, alors

$$F_X(x) = \sum_{t \in D: t \leq x} \mathbb{P}(X = t),$$

avec la convention que la somme est nulle si l'ensemble $\{t \in D : t \leq x\}$ est vide.

Exemple 1.3.5 Si X est une variable aléatoire continue de densité f , alors

$$F_X(x) = \int_{-\infty}^x f(t)dt$$

pour tout $x \in \mathbb{R}$. En particulier, on remarque que si la densité f est continue, alors la fonction F_X est de classe C^1 et $F'_X = f$.

Les propriétés des fonctions de répartition sont rassemblées dans la proposition suivante.

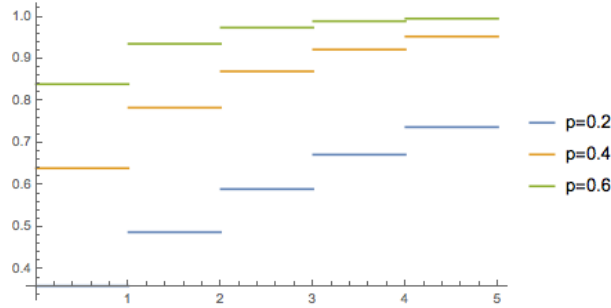


FIGURE 1.2 – Fonction de répartition d’une variable aléatoire géométrique pour $p = 0.2, 0.4$ et 0.6

Proposition 1.3.6 Soit X une variable aléatoire. Alors

- F_X est croissante,
- F_X est continue à droite,
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Démonstration : Cela découle directement des propriétés des mesures rappelées dans la Proposition A.3.8 puisque $F_X(x) = \mathbb{P}(X \leq x) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x + \frac{1}{n})$. ■

Remarque 1.3.7 Si on travaille avec une variable aléatoire sur $\overline{\mathbb{R}}$, alors il est possible d’avoir $\lim_{x \rightarrow +\infty} F_X(x) < 1$ (ce qui arrive lorsque $\mathbb{P}(X = +\infty) > 0$) ou encore $\lim_{x \rightarrow -\infty} F_X(x) > 0$ (lorsque $\mathbb{P}(X = -\infty) > 0$).

Inversement, si on se donne une fonction $F : \mathbb{R} \rightarrow [0, 1]$ satisfaisant les conditions de la Proposition 1.3.6, alors il existe une unique mesure de probabilité μ telle que $\mu(]-\infty, x]) = F(x)$ pour tout $x \in \mathbb{R}$: c’est la mesure de Lebesgue-Stieltje rappelée dans le Théorème A.4.6. Ainsi, on peut interpréter F comme la fonction de répartition d’une variable aléatoire et on dit donc que F est une fonction de répartition.

Exemple 1.3.8 La fonction $F = (1 - p)\mathbb{1}_{[0,1[} + \mathbb{1}_{[1,+\infty[}$ est la fonction de répartition de la variable aléatoire qui prend les valeurs 0 et 1 avec probabilités $1 - p$ et p respectivement.

Exemple 1.3.9 La fonction F illustrée en Figure 1.2 est la fonction de répartition d’une variable de loi géométrique, c’est-à-dire une variable dont la fonction de masse est donnée par

$$\mathbb{P}(X = n) = p(1 - p)^{n-1}, n \in \mathbb{N}_0.$$

Exemple 1.3.10 Les fonctions $F_X(x) = \frac{1}{1+e^{-x}}$, $x \in \mathbb{R}$ et $F_Y(x) = \begin{cases} \frac{1-\epsilon}{1+e^{-x}} & \text{si } x < 0 \\ \epsilon + \frac{1-\epsilon}{1+e^{-x}} & \text{si } x \geq 0 \end{cases}$ sont des fonctions de répartition, cfr Figure 1.3. La première correspond à une expérience dite *logistique* ; la seconde pourrait être une expérience logistique avec un problème en 0.

Les probabilités classiques liées à X peuvent se lire sur sa fonction de répartition, comme illustré sur la Figure 1.4.

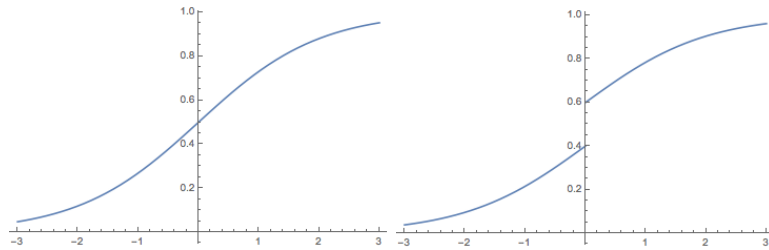


FIGURE 1.3 – Fonction de répartition de X (à gauche) et de Y avec $\epsilon = 0.2$ (à droite)

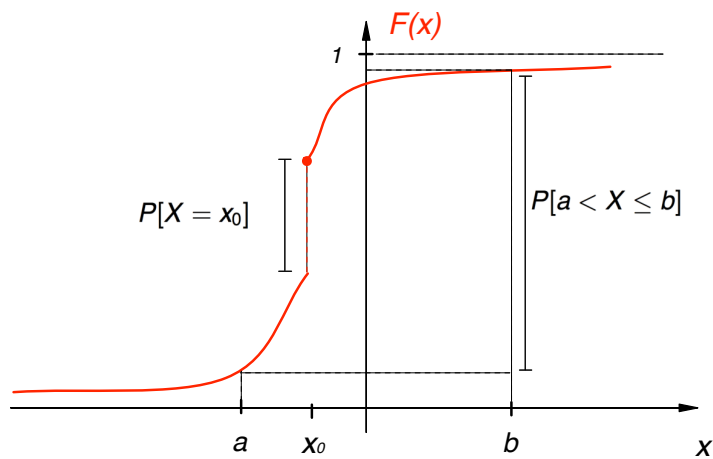


FIGURE 1.4 – Fonction de répartition d'une v.a. X et lecture des probabilités

Théorème 1.3.11 Soit X une variable aléatoire. Pour tous réels $a < b$, on a

- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
- $\mathbb{P}(X > a) = 1 - F_X(a)$
- $\mathbb{P}(X < a) = F_X(a^-)$
- $\mathbb{P}(X = a) = F_X(a) - F_X(a^-)$

où $F_X(a^-) = \lim_{x \rightarrow a^-} F_X(x)$. En particulier, F_X est continue en un point $a \in \mathbb{R}$ si et seulement si $\mathbb{P}(X = a) = 0$.

Démonstration : Les deux premiers points sont évidents. Etant donné que la fonction F_X est croissante et bornée, elle admet une limite à gauche en tout point. Soit $(x_k)_{k \in \mathbb{N}}$ une suite strictement croissante qui converge vers x . Alors $\cup_{k \in \mathbb{N}} \{X \leq x_k\} = \{X < x\}$ et

$$F_X(x^-) = \lim_{k \rightarrow +\infty} F_X(x_k) = \lim_{k \rightarrow +\infty} \mathbb{P}(X \leq x_k) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} \{X \leq x_k\}\right) = \mathbb{P}(X < x),$$

ce qui démontre le troisième point. Pour le quatrième, il suffit de remarquer qu'on peut écrire $\{X \leq x\} = \{X < x\} \cup \{X = x\}$ et d'appliquer ensuite les règles des probabilités ainsi que l'égalité que nous venons de démontrer. ■

Remarque 1.3.12 Ainsi, une fonction de répartition n'est pas nécessairement continue à gauche ; les points x_0 en lesquels $\mathbb{P}(X = x_0) > 0$ sont appelés *atomes* de la loi. La fonction de répartition F_X d'une variable aléatoire est continue sur \mathbb{R} si et seulement si $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$. Dans ce cas, on dit que \mathbb{P}_X est une mesure *diffuse* (qui ne possède pas d'atome). Attention, cela n'implique pas nécessairement que X est une variable aléatoire continue : la loi de Cantor, dont la fonction de répartition est donnée par l'escalier du diable, en est un exemple.

Le résultat suivant montre que l'ensemble des atomes est au plus dénombrable.

Proposition 1.3.13 Si F est une fonction de répartition, alors F admet au plus un nombre dénombrable de points de discontinuité. En particulier, si X est une variable aléatoire, il existe au plus un nombre dénombrable de points $x \in \mathbb{R}$ tels que $\mathbb{P}(X = x) > 0$.

Démonstration : Pour tout $n \in \mathbb{N}_0$, on considère l'ensemble

$$D_n = \left\{ x \in \mathbb{R} : F(x) - F(x^-) \geq \frac{1}{n} \right\}.$$

Puisque F est croissante et à valeurs dans $[0, 1]$, on doit avoir $\#D_n \leq n$. L'ensemble des points de discontinuité $\bigcup_{n \in \mathbb{N}_0} D_n$ de F est donc dénombrable. ■

Corollaire 1.3.14 Soit X une variable aléatoire.

- Si X est discrète, alors F_X est une fonction en escalier possédant un nombre fini ou dénombrable de sauts. Ces sauts ont lieu en les points $x \in \mathbb{R}$ tels que $\mathbb{P}(X = x) > 0$ et la taille du saut y est donnée par $\mathbb{P}(X = x)$.

- Si X est continue de densité f , alors F_X est une fonction continue sur \mathbb{R} . De plus, si f est continue en x , alors F_X est dérivable en x et $F'_X(x) = f(x)$.

Démonstration : Le premier point est évident vu les résultats précédents. Pour le deuxième, il suffit de démontrer la seconde partie. Supposons donc que f est continue en x . Par conséquent, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$|f(y) - f(x)| \leq \varepsilon \quad \text{si} \quad |y - x| \leq \delta.$$

Si $0 < h \leq \delta$, on a

$$\frac{F_X(x+h) - F_X(x)}{h} - f(x) = \frac{1}{h} \int_x^{x+h} f(y) dy - f(x) = \frac{1}{h} \int_x^{x+h} (f(y) - f(x)) dy$$

d'où

$$\left| \frac{F_X(x+h) - F_X(x)}{h} - f(x) \right| \leq \frac{1}{h} \int_x^{x+h} |f(y) - f(x)| dy \leq \varepsilon.$$

On procède de même lorsque $-\delta < h < 0$ et on obtient la conclusion. ■

Remarque 1.3.15 Une variable aléatoire est continue si et seulement si sa fonction de répartition est absolument continue sur \mathbb{R} . Une fonction F est absolument continue sur $[a, b]$ s'il existe une fonction intégrable f sur $[a, b]$ telle que $F(x) = F(a) + \int_a^x f(t) dt$ pour tout $x \in [a, b]$. La notion d'absolue continuité est quelque part entre la notion de continuité uniforme et celle de différentiabilité, l'exemple archétypal étant celui de la fonction valeur absolue. Toute fonction absolument continue est dérivable presque partout. Si la fonction F est absolument continue, on dit que f est sa dérivée. Remarquons que si F n'est pas dérivable partout alors f n'est pas définie de manière unique : en les points $x \in \mathbb{R}$ où F' n'existe pas, on attribue une valeur arbitraire à f et ceci n'aura aucune influence sur le calcul des intégrales.

Une application intéressante des fonctions de répartition est donnée par le résultat suivant, qui permet de simuler numériquement une variable aléatoire de fonction de répartition F à partir de la simulation d'une variable aléatoire uniforme sur $[0, 1]$. Cela donne également une preuve alternative à la réciproque de la Proposition 1.3.6 (sans passer par les mesures de Lebesgue-Stieltjes).

La plupart des logiciels de calcul disposent d'un générateur de nombres aléatoirement distribués sur $[0, 1]$. Le théorème de Skorokhod permet de générer des nombres aléatoirement distribués selon n'importe quelle loi à partir de la loi uniforme.

Proposition 1.3.16 Soit F une fonction de répartition strictement croissante continue et U une v.a. de loi uniforme $U \sim \text{Unif}([0, 1])$. Alors $F^{-1}(U) \sim F$.

Démonstration : Si F est strictement croissante et continue alors elle est bijective et donc inversible. On calcule

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$$

donc la variable aléatoire $Y = F^{-1}(U)$ admet F comme fonction de répartition. ■

La proposition 1.3.16 nous enseigne que pour *toute* fonction de répartition F strictement croissante continue, il est possible de construire une expérience aléatoire et une variable aléatoire explicitement de façon à ce que F en soit la fonction de répartition. Cette vérité ne se limite pas au cas strictement croissant continu.

Théorème 1.3.17 (Représentation de Skorokhod) Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction de répartition. Pour tout $\omega \in [0, 1]$, on pose

$$X^+(\omega) =: \inf \{z : F(z) > \omega\} (= \sup \{y : F(y) \leq \omega\}) \quad (1.2)$$

et

$$X^-(\omega) =: \inf \{z : F(z) \geq \omega\} (= \sup \{y : F(y) < \omega\}). \quad (1.3)$$

Alors, les fonctions X^- et X^+ sont des variables aléatoires réelles définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}[0, 1], \lambda_{|[0,1]})$ de loi F et telles que

$$\mathbb{P}(X^+ = X^-) = 1. \quad (1.4)$$

Démonstration : Par définition de X^- , on a

$$\{\omega \leq F(c)\} \subseteq \{X^-(\omega) \leq c\}. \quad (1.5)$$

De même

$$\{z > X^-(\omega)\} \subseteq \{F(z) \geq \omega\}$$

donc, par continuité à droite de F , on a $F(X^-(\omega)) \geq \omega$ et

$$\{X^-(\omega) \leq c\} \subseteq \{\omega \leq F(X^-(\omega)) \leq F(c)\}. \quad (1.6)$$

Donc, par (1.5) et (1.6) on déduit

$$\{\omega \leq F(c)\} \subseteq \{X^-(\omega) \leq c\}, \quad (1.7)$$

de sorte que

$$\mathbb{P}(X^- \leq c) = F(c).$$

Par définition de X^+ , on a

$$\{\omega < F(c)\} \subseteq \{X^+(\omega) \leq c\},$$

de sorte que $F(c) \leq \mathbb{P}(X^+ \leq c)$. Puisque $X^- \leq X^+$, il est clair que

$$\{X^- \neq X^+\} = \bigcup_{c \in \mathbb{Q}} \{X^- \leq c < X^+\}.$$

Mais, pour tout $c \in \mathbb{R}$,

$$\mathbb{P}(X^- \leq c < X^+) = \mathbb{P}(\{X^- \leq c\} \setminus \{X^+ \leq c\}) \leq F(c) - F(c) = 0.$$

Etant donné que \mathbb{Q} est dénombrable, le résultat suit. ■

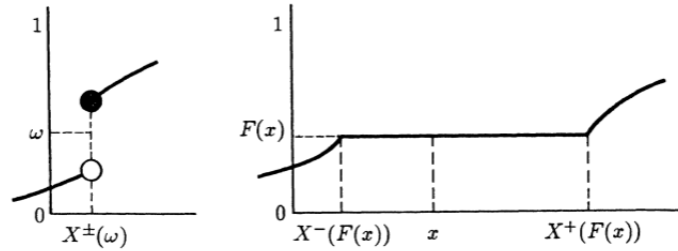


FIGURE 1.5 – Illustration des inverses généralisés X^+ et X^- .

Remarque 1.3.18 Le théorème de représentation ?? nous permet de modéliser les expériences aléatoires directement en travaillant sur les fonctions de répartition *sans avoir à gérer le problème complexe de construction d’une mesure*. Nous dirons simplement “Soit X de fonction de répartition F ” (et on écrit $X \sim F$) sans jamais préciser l’espace sur lequel X est défini.

Terminons cette section en introduisant la fonction quantile.

Définition 1.3.19 Soit X une variable aléatoire de fonction de répartition F .

- On définit sa *fonction quantile*

$$F^{-1} : [0, 1] \rightarrow \mathbb{R} : \alpha \mapsto F^{-1}(\alpha) := \inf \{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

C’est l’inverse généralisé de F .

- Le *quantile d’ordre α* de X est le nombre

$$x_\alpha := F^{-1}(\alpha).$$

- La *médiane* de X est le quantile d’ordre $\frac{1}{2}$.
- Les *quartiles* sont les quantiles d’ordre $\frac{1}{4}$ et $\frac{3}{4}$.

1.4 Lois usuelles

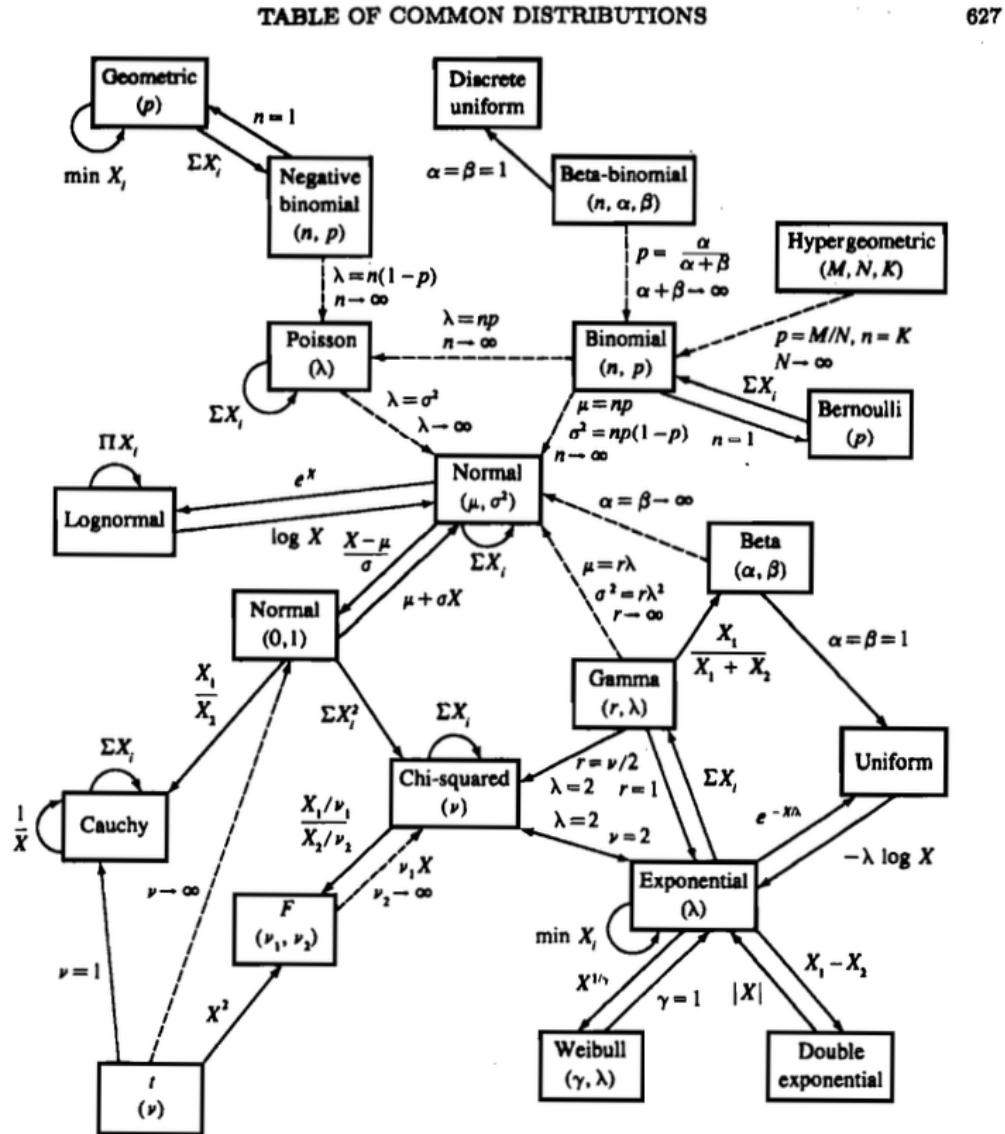
Grâce aux résultats précédemment obtenus, nous savons que *toute* suite de nombres positifs sommant à 1, *toute* fonction positive intégrable d’intégrale 1, *toute* fonction de répartition peuvent être représentés comme étant *la loi* d’une variable aléatoire. Il y a donc une diversité infinie dans les lois de probabilité que l’on peut ainsi obtenir. Seules certaines – les plus utiles – se voient attribuer un nom ; cela nous fait encore une faune gigantesque à explorer (cf Figure 1.6). Nous en parcourons les plus célèbres.

1.4.1 Distributions univariées discrètes

Loi uniforme discrète

Une variable aléatoire X est de *loi uniforme* sur $\{1, \dots, n\}$ si $X(\Omega) = \{1, \dots, n\}$ avec

$$\mathbb{P}(X = k) = \frac{1}{n} \quad \text{pour tout } k \in \{1, \dots, n\}$$



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

FIGURE 1.6 – Capture d'écran du livre de [3]

c'est-à-dire

$$\mathbb{P}_X = \sum_{k=1}^n \frac{1}{n} \delta_k.$$

On note

$$X \sim \text{Unif} \{1, \dots, n\}.$$

Cette loi permet de modéliser des expériences dont les issues sont équiprobables. On peut définir une loi uniforme discrète sur n'importe quel ensemble fini mais il est évidemment impossible de définir une expérience uniforme sur un ensemble infini.

Exemple 1.4.1

- Résultat d'un dé : $X \sim \text{Unif}(\{1, \dots, 6\})$.
- Nombre choisi au hasard entre 1 et 100 : $X \sim \text{Unif}(\{1, \dots, 100\})$.

Loi de Bernoulli

Une variable aléatoire X est de *loi Bernoulli* de paramètre $p \in]0, 1[$ si $X(\Omega) = \{0, 1\}$ avec

$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p$$

ce qu'on écrit $\mathbb{P}(X = x) = p^x(1 - p)^{1-x}$. On a donc

$$\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0$$

et on note

$$X \sim \text{Bern}(p).$$

Cette loi permet de modéliser des expériences n'ayant que deux issues possibles : le succès avec une probabilité p et l'échec avec une probabilité $1 - p$. On code le succès par 1 et l'échec par 0.

Exemple 1.4.2

- Lancer d'une pièce : $X = 1$ si "pile" et $X = 0$ si "face". Alors $X \sim \text{Bern}(1/2)$.
- On veut faire 6 lors du lancer d'un dé : $X = 1$ si le résultat est 6, $X = 0$ sinon. Alors $X \sim \text{Bern}(1/6)$.

Loi Binomiale

Une variable aléatoire X est de *loi binomiale* de paramètres $(n, p) \in \mathbb{N} \times]0, 1[$ si $X(\Omega) = \{0, 1, \dots, n\}$ avec

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k} \quad \text{pour tout } k \in \{0, \dots, n\}.$$

On a donc

$$\mathbb{P}_X = \sum_{k=0}^n C_n^k p^k (1 - p)^{n-k} \delta_k$$

et on note

$$X \sim \text{Bin}(n, p).$$

Cette loi permet de modéliser le nombre de succès lors de n répétitions indépendantes d'une même expérience de type Bernoulli.

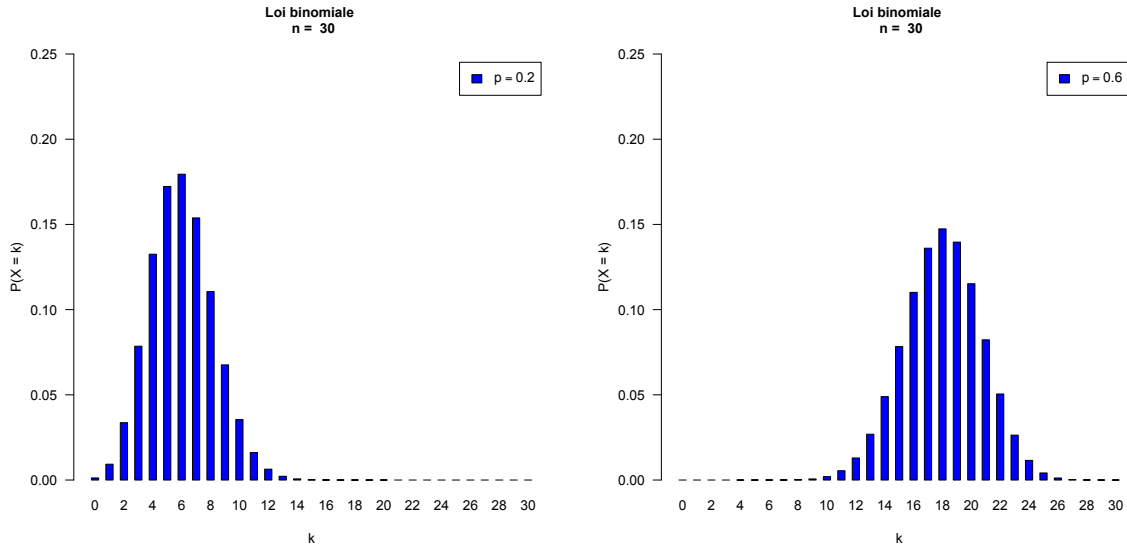


FIGURE 1.7 – Diagrammes en bâtons de lois binomiales de paramètres $(30, 0.2)$ et $(30, 0.6)$.

Exemple 1.4.3

- Nombre de “pile” obtenus lors du lancer de 10 pièces : $X \sim \text{Bin}(10, 1/2)$.
- Nombre d’étudiants malades dans un amphî de 100 étudiants avec la probabilité d’être malade égale à 0.6 : $X \sim \text{Bin}(100, 0.6)$.

Loi Géométrique

Une variable aléatoire X est de *loi géométrique* de paramètre $p \in]0, 1[$ si $X(\Omega) = \mathbb{N}_0$ avec

$$\mathbb{P}(X = k) = (1 - p)^{k-1}p \quad \text{pour tout } k \in \mathbb{N}_0.$$

On a donc

$$\mathbb{P}_X = \sum_{k \in \mathbb{N}_0} (1 - p)^{k-1}p \delta_k$$

et on note

$$X \sim \text{Geom}(p).$$

La loi géométrique permet de modéliser le nombre d’essais nécessaires avant un premier succès lors de répétitions indépendantes d’une même expérience de type Bernoulli.

Exemple 1.4.4 Nombre de lancers nécessaires pour obtenir “pile” : $X \sim \text{Geom}(1/2)$.

Remarque 1.4.5 Il existe deux paramétrisations classiques de la loi géométrique. La seconde prend ses valeurs dans \mathbb{N} et sa fonction de masse est $\mathbb{P}(X = k) = (1 - p)^k p$. Elle modélise le temps du dernier échec.

Proposition 1.4.6 Soit X une variable aléatoire.

1. $X \sim \text{Geom}(p)$ si et seulement si $\mathbb{P}(X > k) = (1 - p)^k$ pour tout $k \in \mathbb{N}$.
2. La distribution géométrique est “sans mémoire” : si $X \sim \text{Geom}(p)$, alors pour tous $j, k \geq 1$,

$$\mathbb{P}(X \geq k + j | X > k) = \mathbb{P}(X \geq j).$$

Démonstration : Pour le premier point, il suffit de remarquer que si $X \sim \text{Geom}(p)$, alors

$$\mathbb{P}(X > k) = \sum_{j=k+1}^{+\infty} \mathbb{P}(X = j) = \sum_{j=k+1}^{+\infty} (1 - p)^{j-1} p = (1 - p)^k.$$

La réciproque est immédiate puisque la fonction de répartition caractérise la loi. Pour le second point, on utilise la définition des probabilités conditionnelles pour écrire

$$\mathbb{P}(X \geq k + j | X > k) = \frac{\mathbb{P}(X \geq k + j, X > k)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X \geq k + j)}{\mathbb{P}(X > k)}$$

puisque $\{X \geq j + k\} \subseteq \{X > k\}$. Par le premier point et puisque X est à valeurs entières, on trouve que

$$\frac{\mathbb{P}(X \geq k + j)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X > k + j - 1)}{\mathbb{P}(X > k)} = \frac{(1 - p)^{k+j-1}}{(1 - p)^k} = (1 - p)^{j-1}.$$

En appliquant à nouveau le premier point, on obtient

$$(1 - p)^{j-1} = \mathbb{P}(X > j - 1) = \mathbb{P}(X \geq j),$$

ce qui donne le résultat attendu. ■

Loi binomiale négative (distribution de Pascal)

Une variable aléatoire X est de *loi binomiale négative* de paramètres $(r, p) \in \mathbb{N}_0 \times [0, 1]$ si $X(\Omega) = \{r, r + 1, \dots\}$ avec

$$\mathbb{P}(X = k) = C_{k+r-1}^k (1 - p)^r p^k \quad \text{pour tout } k \in \{r, r + 1, \dots\}$$

c'est-à-dire

$$\mathbb{P}_X = \sum_{k=r}^{+\infty} C_{k+r-1}^k (1 - p)^r p^k \delta_k.$$

On note

$$X \sim \text{NB}(r, p).$$

La loi binomiale négative permet de modéliser le nombre de succès avant r échecs lors de répétitions d'une expérience de type Bernoulli avec probabilité de succès p . Le cas $r = 1$ correspond à la loi géométrique (où les rôles de succès et d'échec sont inversés).

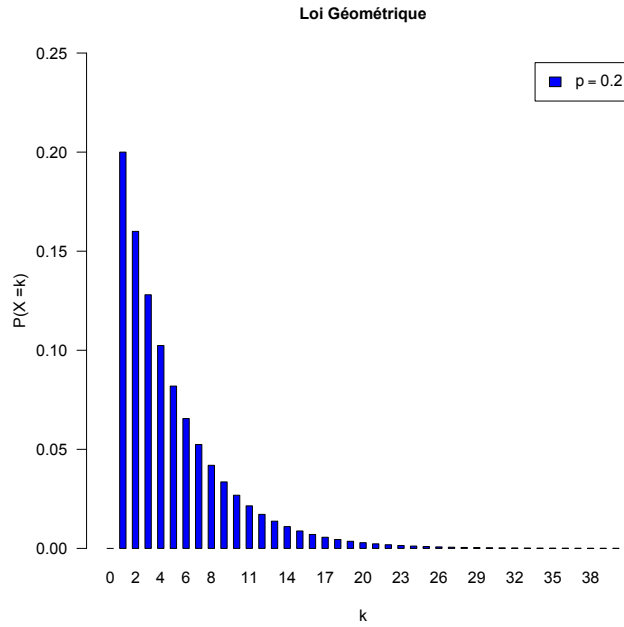


FIGURE 1.8 – Diagramme en bâtons d’une loi géométrique de paramètre 0.2.

Loi de Poisson

Une variable aléatoire X est de *loi Poisson* de paramètre $\lambda > 0$ si $X(\Omega) = \mathbb{N}$ avec

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{pour tout } k \in \mathbb{N}.$$

On a donc

$$\mathbb{P}_X = \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$$

et on note

$$X \sim \text{Poi}(\lambda).$$

La loi de Poisson permet de modéliser le nombre de réalisations d’événements rares.

Exemple 1.4.7

- Nombre d’accidents sur la route par jour.
- Nombre de clients par heure dans un supermarché.
- Nombre de visites sur Wikipedia par unité de temps.

1.4.2 Distributions univariées continues

Loi uniforme

Une variable aléatoire X est de *loi uniforme* sur l’intervalle $[a, b]$ si elle est continue de densité

$$f = \frac{1}{b - a} \mathbb{1}_{[a,b]}.$$

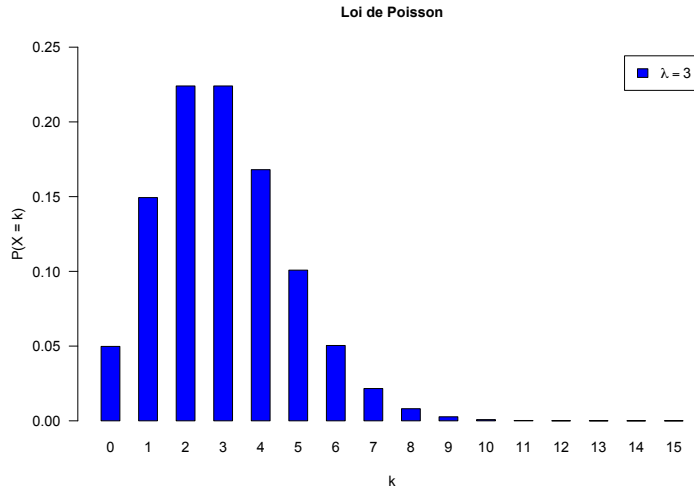


FIGURE 1.9 – Diagramme en bâtons d’une loi de Poisson de paramètre 3.

On note

$$X \sim \text{Unif}(a, b).$$

Cette loi permet de modéliser un choix parfaitement aléatoire d’un réel entre deux bornes a et b .

Loi exponentielle

Une variable aléatoire X est de *loi exponentielle* de paramètre $\lambda > 0$ si elle est continue de densité

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x)$$

pour tout $x \in \mathbb{R}$. On note

$$X \sim \text{Exp}(\lambda).$$

Cette loi permet de modéliser une durée de vie ou un temps d’attente.

Proposition 1.4.8 Soit X une variable aléatoire.

- $X \sim \text{Exp}(\lambda)$ si et seulement si $\mathbb{P}(X \geq t) = \mathbb{P}(X > t) = e^{-\lambda t}$ pour tout $t \geq 0$.
- la distribution exponentielle est “sans mémoire” : si $X \sim \text{Exp}(\lambda)$, alors

$$\mathbb{P}(X \geq t + s \mid X > t) = \mathbb{P}(X \geq s)$$

pour tous $t, s \geq 0$.

Démonstration : Pour le premier point, on calcule directement que si $X \sim \text{Exp}(\lambda)$, alors

$$\mathbb{P}(X \geq t) = \mathbb{P}(X > t) = \int_t^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_t^{+\infty} = e^{-\lambda t}.$$

Pour le deuxième point, on a

$$\begin{aligned} \mathbb{P}(X \geq t + s | X > t) &= \frac{\mathbb{P}(X \geq t + s, X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X \geq t + s)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= \mathbb{P}(X \geq s). \end{aligned}$$

■

Cette propriété nous informe que la probabilité qu'un composant survive t unités de temps supplémentaires ne dépend pas de l'âge du composant. La loi exponentielle et la loi géométrique sont des versions continue et discrète d'un même problème de modélisation.

Proposition 1.4.9 Si $X \sim \text{Exp}(p)$ et $Y = \lceil X \rceil$ alors Y est de loi géométrique de paramètre $1 - e^{-\lambda}$.

Démonstration : Bien entendu, Y est à valeurs dans \mathbb{N}_0 . De plus, pour tout $k \geq 1$, on a

$$\begin{aligned} \mathbb{P}(Y = k) &= \mathbb{P}(X \in]k - 1, k]) = \int_{k-1}^k \lambda e^{-\lambda x} dx \\ &= (e^{-\lambda})^{k-1} (1 - e^{-\lambda}), \end{aligned}$$

ce qui suffit par la Proposition 1.4.6. ■

Loi gamma

Une variable aléatoire X est de *loi gamma* de paramètres $(\alpha, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+$ si elle est continue de densité

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{[0, +\infty[}(x)$$

pour tout $x \in \mathbb{R}$, où $\Gamma(t) = \int_0^{+\infty} e^{-x} x^{t-1} dx$. On note

$$X \sim \Gamma(\alpha, \beta).$$

Si $\alpha \in \mathbb{N}_0$ alors on parle de loi *Erlang*.

La loi gamma permet de modéliser une durée de vie (ou un temps d'attente) d'un composant composé de plusieurs éléments indépendants.

Loi Gaussienne standard


Une variable aléatoire X est de *distribution normale standard* (ou *Gaussienne standard*) si elle est continue de densité

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

pour tout $x \in \mathbb{R}$. On note

$$X \sim \mathcal{N}(0, 1).$$

Table of the Normal Distribution



**Probability Content
from $-\infty$ to Z**

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

FIGURE 1.10 – Table de la fonction de répartition de la loi $\mathcal{N}(0, 1)$

Cette loi apparaît partout (la plus importante de toutes les lois!); elle permet notamment de modéliser les erreurs de mesure.

La fonction de répartition n’admet pas de forme fermée; on la note

$$\Phi(x) = \int_{-\infty}^x \varphi(u)du. \tag{1.8}$$

Cette fonction n’est calculable que par intégration numérique. On utilisera les *tables de la loi normale* (cf Figure 1.10) ou un logiciel de calcul.

On utilise énormément les propriétés de symétrie des lois normales, qui découlent toutes de l’identité $\varphi(x) = \varphi(-x)$ pour tout $x \in \mathbb{R}$.

Proposition 1.4.10 Si $X \sim \mathcal{N}(0, 1)$ alors

- $\mathbb{P}(X < -x) = \mathbb{P}(X > x)$ (et en particulier $\mathbb{P}(X > 0) = \mathbb{P}(X < 0) = \frac{1}{2}$).
- $\Phi(-x) = 1 - \Phi(x)$ (et donc on ne tabule que $\Phi(x), x \geq 0$).
- $\mathbb{P}(|X| > x) = 2\mathbb{P}(X > x)$.

Loi Gaussienne

Une variable aléatoire X est de *distribution normale* de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}_0^+$ si X admet la représentation en loi

$$X \stackrel{\mathcal{L}}{=} \sigma Z + \mu \quad \text{avec} \quad Z \sim \mathcal{N}(0, 1).$$

On vérifie que X est continue de densité

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

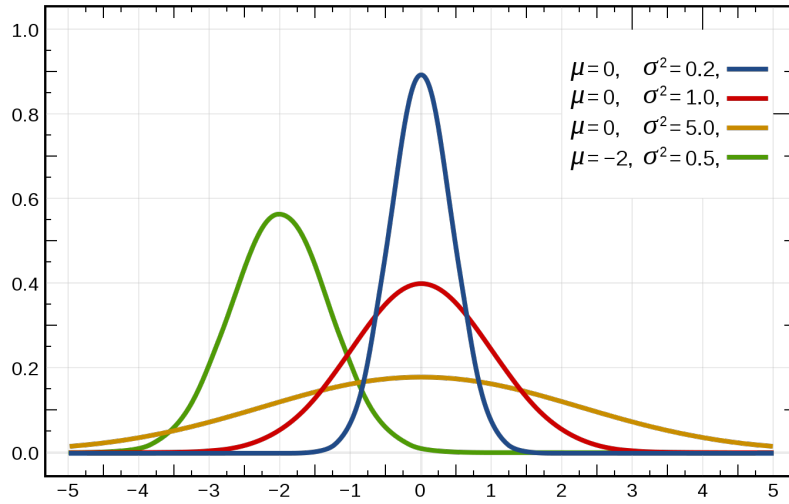


FIGURE 1.11 – Quelques représentations de densités de lois normales pour des paramètres différents. On a une courbe en “cloche”, symétrique par rapport à l’axe $x = \mu$. Plus σ est grand, plus la courbe est “aplatie”.

pour tout $x \in \mathbb{R}$. On note

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Définition 1.4.11 Centrer-réduire une variable $X \sim \mathcal{N}(\mu, \sigma^2)$ est équivalent à la remplacer par la variable aléatoire $Z = (X - \mu)/\sigma$ qui, par définition, est de loi $\mathcal{N}(0, 1)$. On dit aussi parfois qu’on *standardise* X .

Deux lois dérivées de la Gaussienne

Les deux lois présentées ci-dessous font intervenir la notion d’indépendance de variables aléatoires. Nous y revenons plus tard.

1. **Loi χ^2 à ν degrés de liberté** : $X \sim \chi_\nu^2$ si

$$X = \sum_{i=1}^{\nu} Z_i^2$$

où les Z_i sont des $\mathcal{N}(0, 1)$ indépendantes¹. Dans ce cas, on peut montrer que

$$f_X(x) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} e^{-\frac{x}{2}} x^{(\nu-2)/2} \mathbb{1}_{[0, +\infty[}(x).$$

2. **Loi de Student à ν degrés de liberté** : $X \sim t_\nu$ si

$$X = \frac{Z}{\sqrt{Y/\nu}}$$

avec $Z \sim \mathcal{N}(0, 1)$ indépendant de $Y \sim \chi_\nu^2$. Dans ce cas, on peut montrer que

$$f_X(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

1. Nous introduirons cette notion plus tard.

Ces deux lois sont – même si cela peut sembler étrange – très naturelles. Vous y reviendrez dans votre cours de statistiques.

1.5 Vecteurs aléatoires

Focalisons-nous maintenant sur les variables aléatoires à image dans $(\mathbb{R}^d, \mathcal{B}^d)$ pour un entier d supérieur ou égal à 1.

Définition 1.5.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Un *vecteur aléatoire* est une application mesurable $X : \Omega \rightarrow \mathbb{R}^d$, i.e. telle que

$$X^{-1}(B) \in \mathcal{F} \text{ pour tout } B \in \mathcal{B}^d.$$

Un vecteur aléatoire de \mathbb{R}^d n'est rien d'autre qu'un d -uple de variables aléatoires réelles puisque $\mathcal{B}^d = \mathcal{B} \times \dots \times \mathcal{B}$. Ainsi, $X : \Omega \rightarrow \mathbb{R}^d$ est un vecteur aléatoire si et seulement si ses composantes sont des variables aléatoires.

Remarque 1.5.2 La convention veut que, lorsque c'est pertinent, on prend le vecteur en colonne plutôt qu'en ligne. Cette convention jouera un rôle dans votre cours de statistique multivariée.

Comme dans le cas univarié, on écrira "soit X un vecteur aléatoire" sans spécifier l'espace de départ $(\Omega, \mathcal{F}, \mathbb{P})$. Comme précédemment, ce ne sont pas tant les valeurs spécifiques prises par X qui nous intéressent, mais plutôt la fréquence potentielle des occurrences de X dans les différentes parties de l'espace.

Remarque 1.5.3 Pour toute mesure de probabilité μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire $X : \Omega \rightarrow \mathbb{R}^d$ telle que $\mathbb{P}_X = \mu$: il suffit comme dans le cas univarié de prendre $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}^d$, $\mathbb{P} = \mu$ et de considérer la variable aléatoire $X = \text{id}$. Cela justifie le fait que l'on parle de vecteur aléatoire X ayant une loi de probabilité \mathbb{P}_X donnée sans spécifier l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel X est défini. En revanche, il est clair qu'il n'y a pas d'unique vecteur aléatoire X vérifiant $\mathbb{P}_X = \mu$: comme pour les variables aléatoires, un vecteur aléatoire n'est donc pas déterminé par sa loi.

Définition 1.5.4 Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. La *loi* de X est la mesure image \mathbb{P}_X de \mathbb{P} par X , i.e. c'est la mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$ définie par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B)$$

pour tout $B \in \mathcal{B}^d$. Si $j \in \{1, \dots, d\}$, la loi de X_j est appelée la *$j^{\text{ème}}$ loi marginale* du vecteur X . La loi \mathbb{P}_X est également appelée *distribution jointe* des variables aléatoires X_1, \dots, X_d .

Si deux vecteurs aléatoires X et Y ont la même loi, on dit qu'ils sont identiquement distribués et on note $X \stackrel{\mathcal{L}}{=} Y$. Comme dans le cas univarié, on peut introduire la fonction de répartition d'un vecteur aléatoire.

Définition 1.5.5 Soit X un vecteur aléatoire. La *fonction de répartition* F_X de X est définie

par

$$\begin{aligned}
 F_X : \mathbb{R}^d \rightarrow [0, 1] : t = (t_1, \dots, t_d) \mapsto F_X(t) &= \mathbb{P}_X (]-\infty, t_1] \times \dots \times]-\infty, t_d]) \\
 &= \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d).
 \end{aligned}$$

Pour les mêmes raisons qu'en dimension 1, puisque les demi-plans engendrent la σ -algèbre de Borel, la donnée de la distribution jointe d'un vecteur aléatoire est parfaitement équivalente à la donnée de sa fonction de répartition.

Proposition 1.5.6 *Deux vecteurs aléatoires X et Y sont identiquement distribués si et seulement si $F_X = F_Y$.*

Par continuité des mesures de probabilité, les fonctions de répartition marginales s'obtiennent en passant à la limite :

$$F_{X_i}(t_i) = \mathbb{P}(X_i \leq t_i) = \lim_{t_j \rightarrow +\infty, j \neq i} F_X(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_d).$$

La donnée de la loi du vecteur permet de déduire celle des marginales; la réciproque est généralement fausse.

Exemple 1.5.7 Considérons l'expérience consistant à lancer deux dés (distinguables). Alors $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$. Le vecteur aléatoire (X_1, X_2) rendant compte du résultat obtenu est

$$(X_1, X_2) : \Omega \rightarrow \mathbb{R}^2 : (i, j) \mapsto (X_1, X_2)(i, j) = (i, j)$$

et a une loi jointe triviale à calculer donnée dans la table 1.2. Les lois marginales s'obtiennent

	1	2	3	4	5	6	
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

TABLE 1.1 – Loi jointe et lois marginales des résultats du lancer de deux dés distinguables

en sommant les probabilités en ligne ou en colonne de la table de la loi jointe. On notera que les entrées de la loi jointe s'obtiennent en prenant le produit des entrées marginales. On peut écrire la loi sous la forme $\mathbb{P}_{(X_1, X_2)} = \frac{1}{36} \sum_{i,j=1}^6 \delta_{(i,j)}$. On peut aisément construire un autre vecteur

aléatoire bivarié possédant les mêmes entrées marginales. Par exemple, soit X le résultat obtenu en lançant un dé une seule fois. On prend $X_1 = X$ et $X_2 = 7 - X_1$. Alors le couple (X_1, X_2) a la loi jointe donnée dans le tableau 1.2.

	1	2	3	4	5	6	
1	0	0	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$
2	0	0	0	0	$\frac{1}{6}$	0	$\frac{1}{6}$
3	0	0	0	$\frac{1}{6}$	0	0	$\frac{1}{6}$
4	0	0	$\frac{1}{6}$	0	0	0	$\frac{1}{6}$
5	0	$\frac{1}{6}$	0	0	0	0	$\frac{1}{6}$
6	$\frac{1}{6}$	0	0	0	0	0	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

TABLE 1.2 – Loi jointe et lois marginales du couple $(X, 7 - X)$ pour X le résultat d’un lancer

Exemple 1.5.8 Dans la suite de l’exemple précédent, notons maintenant X_1 la somme des résultats de chaque dé et X_2 la différence des résultats de chaque dé (en valeur absolue) :

$$(X_1, X_2) : \Omega \rightarrow \mathbb{R}^2 : (i, j) \mapsto (X_1, X_2)(i, j) = (i + j, |i - j|).$$

Séparément, X_1 et X_2 définissent chacun une variable aléatoire ; ensemble ils forment un vecteur aléatoire. La loi jointe et les lois marginales sont résumées dans le tableau 4.1. Les lois marginales

	2	3	4	5	6	7	8	9	10	11	12	
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{6}{36}$
1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{10}{36}$
2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	$\frac{8}{36}$
3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	$\frac{6}{36}$
4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0	$\frac{4}{36}$
5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0	$\frac{2}{36}$
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

TABLE 1.3 – Loi jointe et lois marginales de la somme et la différence des résultats du lancer de deux dés distinguables

s’obtiennent toujours en sommant les probabilités en ligne ou en colonne de la table de la loi

jointe, mais la loi jointe ne s'obtient pas comme produit des entrées marginales correspondantes. On peut écrire la loi sous la forme

$$\mathbb{P}_{(X_1, X_2)} = \sum_{i=2}^{12} \sum_{j=0}^5 p_{ij} \delta_{(i,j)}$$

où p_{ij} est l'entrée (i, j) du tableau de la loi jointe.

Les exemples précédents sont tous discrets. Comme dans le cas univarié, on distingue deux cas particuliers de vecteurs aléatoires.

Définition 1.5.9 Soit X un vecteur aléatoire dans \mathbb{R}^d , $d \geq 1$. On dit que

- le vecteur aléatoire X est *discret* si sa loi est discrète, c'est-à-dire s'il existe une suite $(p_{i_1, \dots, i_d})_{i_1, \dots, i_d}$ de réels positifs ou nuls et une suite $(x_{1, i_1}, \dots, x_{d, i_d})_{i_1, \dots, i_d}$ de vecteurs de \mathbb{R}^d tels que

$$\mathbb{P}_X = \sum_{i_1} \dots \sum_{i_d} p_{i_1, \dots, i_d} \delta_{(x_{1, i_1}, \dots, x_{d, i_d})}.$$

- le vecteur aléatoire X est *continu* si sa loi est absolument continue par rapport à la mesure de Lebesgue dans \mathbb{R}^d , c'est-à-dire s'il existe une fonction mesurable $f : \mathbb{R}^d \rightarrow [0, +\infty]$ telle que

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B f(u) du$$

pour tout $B \in \mathbb{B}^d$. La fonction f est appelée *fonction de densité* de X .

Evidemment, comme dans le cas univarié, un vecteur aléatoire n'est pas discret ou continu ; il existe une palette d'options intermédiaires qu'on ne peut négliger.

Intéressons-nous dans un premier temps aux vecteurs aléatoires discrets. Clairement, puisque $\mathbb{P}_X(\mathbb{R}^d) = 1$, on a

$$\sum_{i_1} \dots \sum_{i_d} p_{i_1, \dots, i_d} = 1.$$

On va se placer dans le cas bivarié ($d = 2$) pour simplifier les notations. Tout se généralise aisément aux dimensions supérieures. Les distributions marginales s'obtiennent par projection.

Proposition 1.5.10 Soit (X_1, X_2) un vecteur aléatoire discret de loi

$$\mathbb{P}_{(X_1, X_2)} = \sum_{i \in I} \sum_{j \in J} p_{ij} \delta_{(x_{1, i}, x_{2, j})}.$$

Alors, les variables aléatoires sont discrètes et de loi

$$\mathbb{P}_{X_1} = \sum_{i \in I} \left(\sum_{j \in J} p_{ij} \right) \delta_{x_{1, i}} \quad \text{et} \quad \mathbb{P}_{X_2} = \sum_{j \in J} \left(\sum_{i \in I} p_{ij} \right) \delta_{x_{2, j}}$$

respectivement.

Démonstration : Traitons le cas de la loi X_1 . Pour tout sous-ensemble borélien B de \mathbb{R} , on a

$$\begin{aligned} \mathbb{P}_{X_1}(B) &= \mathbb{P}(X_1 \in B) = \mathbb{P}(X_1 \in B, X_2 \in \mathbb{R}) \\ &= \sum_{i \in I, j \in J} p_{ij} \delta_{(x_{1,i}, x_{2,j})}(B \times \mathbb{R}) \\ &= \sum_{i \in I} \left(\sum_{j \in J} p_{ij} \right) \delta_{x_{1,i}}(B) \end{aligned}$$

puisque $\delta_{(x_{1,i}, x_{2,j})}(B \times \mathbb{R}) = \delta_{x_{1,i}}(B)$. ■

Soit (X_1, X_2) un vecteur aléatoire bivarié discret. On appelle *fonction de masse* de X la fonction

$$P_{(X_1, X_2)} : (x_1, x_2) \in \mathbb{R}^2 \mapsto \mathbb{P}((X_1, X_2) = (x_1, x_2)).$$

On a

$$\mathbb{P}((X_1, X_2) \in B) = \sum_{i, j: (x_{1,i}, x_{2,j}) \in B} P_{(X_1, X_2)}(x_{1,i}, x_{2,j})$$

pour tout $B \in \mathcal{B}^2$. Le résultat précédent montre que les fonctions de masse marginales s'obtiennent par projection, c'est-à-dire

$$\begin{aligned} P_{X_1}(x_{1,i}) &= \sum_{j \in J} P_{(X_1, X_2)}(x_{1,i}, x_{2,j}) \text{ pour tout } i \in I \\ P_{X_2}(x_{2,j}) &= \sum_{i \in I} P_{(X_1, X_2)}(x_{1,i}, x_{2,j}) \text{ pour tout } j \in J. \end{aligned}$$

Dans le cas continu, on remarque que la fonction de répartition $F_{(X_1, X_2)}$ s'obtient à partir de la fonction de densité f via la relation

$$F_{(X_1, X_2)}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(u, v) dv du$$

et réciproquement

$$f = \frac{d}{dx_1} \frac{d}{dx_2} F_{(X_1, X_2)}$$

en tout point de dérivabilité de la fonction de répartition. On remarque aussi que f est nécessairement une fonction positive telle que

$$\int_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 = 1.$$

Notons que f n'est pas nécessairement une fonction bornée. Les distributions marginales s'obtiennent, exactement comme dans le cas discret, par projection.

Proposition 1.5.11 *Soit $X = (X_1, X_2)$ un vecteur aléatoire continu de densité f . Alors X_1 et X_2 sont des variables aléatoires continues de densité*

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f(x_1, x_2) dx_2 \quad \text{et} \quad f_{X_2}(x_2) = \int_{\mathbb{R}} f(x_1, x_2) dx_1$$

respectivement.

Démonstration : Pour tout ensemble borélien B de \mathbb{R} , on remarque que

$$\begin{aligned} \mathbb{P}_{X_1}(B) = \mathbb{P}(X_1 \in B) &= \mathbb{P}(X_1 \in B, X_2 \in \mathbb{R}) \\ &= \int_{B \times \mathbb{R}} f(x_1, x_2) dx_2 dx_1 \\ &= \int_B \left(\int_{\mathbb{R}} f(x_1, x_2) dx_2 \right) dx_1 \end{aligned}$$

par le théorème de Tonelli-Fubini, d'où la conclusion. ■

Exemple 1.5.12 Une chaîne de restauration rapide vend des hamburgers selon deux modalités distinctes : un comptoir traditionnel et un drive-in. Soit X_1 la proportion du temps où le comptoir traditionnel est occupé le jeudi et X_2 la proportion du temps où le drive-in est occupé le jeudi. Supposons que (X_1, X_2) admette la fonction de densité

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{sinon.} \end{cases}$$

Notez bien qu'il s'agit là d'un choix arbitraire de modélisation, contrairement à l'exemple du lancer de dés où la loi sous-jacente était dictée par l'expérience. Le manager est satisfait si le comptoir traditionnel est occupé au moins la moitié du temps et le drive-in au moins un quart du temps, ce qui arrive avec la probabilité $\mathbb{P}(1/2 \leq X_1 \leq 1, 1/4 \leq X_2 \leq 1)$, laquelle vaut

$$\begin{aligned} \int_{1/2}^1 \int_{1/4}^1 f(x, y) dy dx &= \int_{1/2}^1 \left(\int_{1/4}^1 \frac{6}{5}(x + y^2) dy \right) dx = \int_{1/2}^1 \left[\frac{6}{5} \left(xy + \frac{y^3}{3} \right) \right]_{y=1/4}^{y=1} dx \\ &= \int_{1/2}^1 \left(\frac{9x}{10} + \frac{63}{160} \right) dx = \left[\frac{9x^2}{20} + \frac{63x}{160} \right]_{1/2}^1 = \frac{171}{320}, \end{aligned}$$

ou encore

$$\begin{aligned} \int_{1/4}^1 \int_{1/2}^1 f(x, y) dx dy &= \int_{1/4}^1 \left(\int_{1/2}^1 \frac{6}{5}(x + y^2) dx \right) dy = \int_{1/4}^1 \left[\frac{6}{5} \left(\frac{x^2}{2} + y^2 x \right) \right]_{x=1/2}^{x=1} dy \\ &= \int_{1/4}^1 \left(\frac{9}{20} + \frac{3y^2}{5} \right) dy = \left[\frac{9y}{20} + \frac{y^3}{5} \right]_{1/4}^1 = \frac{171}{320}. \end{aligned}$$

En procédant de la même façon, on obtient

$$F_{(X_1, X_2)}(x, y) = \begin{cases} 0 & \text{si } x < 0 \text{ ou } y < 0 \\ \frac{1}{5}xy(3x + 2y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ \frac{1}{5}x(3x + 2) & \text{si } x \in [0, 1] \text{ et } y > 1 \\ \frac{1}{5}y(3 + 2y^2) & \text{si } x > 1 \text{ et } y \in [0, 1] \\ 1 & \text{sinon.} \end{cases}$$

Pour les marginales, on a

$$f_{X_1}(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{6}{5}(x + y^2) dy = \left[\frac{6xy}{5} + \frac{2y^3}{5} \right]_0^1 = \frac{6x}{5} + \frac{2}{5} \quad \text{si } 0 \leq x \leq 1$$

$$f_{X_2}(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 \frac{6}{5}(x + y^2) dx = \left[\frac{3x^2}{5} + \frac{6xy^2}{5} \right]_0^1 = \frac{3}{5} + \frac{6y^2}{5} \quad \text{si } 0 \leq y \leq 1$$

et ces fonctions prennent la valeur 0 ailleurs.

Nous insistons sur le fait que les lois marginales ne donnent de l'information que sur le comportement univarié des marginales, et n'apportent aucune information sur le comportement joint ! Etant données deux lois marginales X_1, X_2 , on peut imaginer de nombreuses façons de les « coupler » pour former une loi jointe.

Si X est un vecteur aléatoire de dimension d et si $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est une fonction mesurable, alors $h(X)$ est un vecteur aléatoire de dimension d . Le résultat suivant montre que si h est un changement de variables et si X est continu, alors on peut calculer facilement la loi de $h(X)$.

Proposition 1.5.13 *Soit X un vecteur aléatoire de dimension d continu de densité f_X et soit $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ un difféomorphisme. Alors, le vecteur aléatoire $Y = \Phi(X)$ est continu de densité*

$$f_Y(y) = f_X \circ \Phi^{-1}(y) |J_{\Phi^{-1}}(y)|, \quad y \in \mathbb{R}^d$$

où J est le Jacobien du changement de variables.

Démonstration : En utilisant la formule du changement de variables dans \mathbb{R}^d pour Φ^{-1} , on a

$$\mathbb{P}(Y \in B) = \mathbb{P}(X \in \Phi^{-1}(B)) = \int_{\Phi^{-1}(B)} f_X(x) dx = \int_B f_X(\Phi^{-1}(y)) |J_{\Phi^{-1}}(y)| dy,$$

ce qui suffit. ■

Remarque 1.5.14 Ce résultat est évidemment valide lorsque l'on travaille avec des variables aléatoires et des changements de variables dans \mathbb{R} .

Remarque 1.5.15 On peut également appliquer ce résultat dans le cas d'un difféomorphisme $\Phi : U \rightarrow V$ si la variable aléatoire X est à valeurs dans U .

1.6 Lois multivariées usuelles

La situation est exactement la même qu'en dimension 1 : toute fonction positive et intégrable sur \mathbb{R}^d d'intégrale égale à 1, toute suite positive dont la somme est égale à 1, donnent naissance à une loi de probabilité.

1.6.1 Loi multinomiale

Soient $k, n \in \mathbb{N}_0$ et $p_1, \dots, p_k \in [0, 1]$ tels que $\sum_{\ell=1}^k p_\ell = 1$. Un vecteur aléatoire $X = (X_1, \dots, X_k)$ est de *distribution multinomiale* de paramètres n, p_1, \dots, p_k si

$$\mathbb{P}(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{(n_1!) \dots (n_k!)} p_1^{n_1} \dots p_k^{n_k}$$

dès que (n_1, \dots, n_k) sont tels que $\sum_{\ell=1}^k n_\ell = n$. On écrit $X \sim \text{Multin}(n, p_1, \dots, p_k)$.

Elle permet de modéliser des expériences aléatoires à k résultats possibles, avec $\mathbb{P}(\text{résultat } \ell) = p_\ell$.

Exemple 1.6.1 Au premier tour de la dernière élection présidentielle française, on interroge n personnes en leur demandant pour lequel des k candidats elles ont l'intention de voter. En notant X_ℓ le nombre de sondés déclarant vouloir voter pour le candidat ℓ ,

$$(X_1, \dots, X_k)^T \sim \text{Multin}(n, p_1, \dots, p_k),$$

où p_ℓ est la proportion des Français en faveur du candidat ℓ .

Proposition 1.6.2 (Lois marginales) Soit $X \sim \text{Multin}(n, p_1, \dots, p_k)$. Alors

- $X_\ell \sim \text{Bin}(n, p_\ell)$,
- $X_\ell + X_m \sim \text{Bin}(n, p_\ell + p_m)$ si $\ell \neq m$.

1.6.2 Loi normale multivariée

L'extension de la loi gaussienne univariée au cas multivarié se fait typiquement de la façon suivante.

Définition 1.6.3 Un vecteur aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans \mathbb{R}^d est *gaussien* si toute combinaison linéaire de ses composantes est gaussienne.

Si X est un vecteur aléatoire gaussien, alors ses marginales sont également gaussiennes comme le montre le résultat suivant.

Proposition 1.6.4 Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire gaussien, alors X_i suit une loi gaussienne pour tout $i \in \{1, \dots, d\}$.

Démonstration : Si on fixe $i \in \{1, \dots, d\}$, il suffit de considérer la combinaison linéaire $\sum_{j=1}^d \alpha_j X_j$ où $\alpha_i = 1$ et $\alpha_j = 0$ pour tout $j \neq i$. ■

Remarque 1.6.5 L'affirmation réciproque n'est pas vraie : il est possible de construire un vecteur dont toutes les marginales sont gaussiennes sans pour autant qu'elles soient conjointement gaussiennes. En effet, si $Z \sim \mathcal{N}(0, 1)$ et si on pose

$$Z' = \begin{cases} Z & \text{si } |Z| \leq 1 \\ -Z & \text{si } |Z| > 1 \end{cases}$$

alors Z' est également gaussien standard. Mais étant donné que $|Z + Z'| \leq 2$ et $Z + Z'$ n'est pas constante, $Z + Z'$ n'est pas gaussien et donc (Z, Z') n'est pas un vecteur gaussien. La définition 1.6.3 requiert donc strictement plus que la seule normalité des marginales.

La loi normale multivariée est une loi très importante. Nous reviendrons sur ses propriétés plus tard. Signalons néanmoins déjà qu'un vecteur aléatoire gaussien n'a pas une loi discrète mais n'a pas nécessairement une loi continue non plus. Par exemple le vecteur $X = (N, N)$ avec $N \sim \mathcal{N}(0, 1)$ est gaussien mais si D est la première bissectrice, alors $\mathbb{P}(X \in D) = 1$ alors que $\lambda(D) = 0$.

Chapitre 2

Espérance mathématique

La notion d'espérance mathématique joue un rôle central lorsque l'on travaille avec des variables aléatoires ; elle est le coeur même de l'essence de la théorie des probabilités et des statistiques. On pourrait néanmoins résumer l'introduction de cette notion en une phrase : puisque X est une fonction mesurable sur $(\Omega, \mathcal{F}, \mathbb{P})$, l'espérance de X n'est rien d'autre que l'intégrale de cette fonction dans la mesure \mathbb{P} , si celle-ci existe. Toutefois, l'importance de cette notion justifie qu'on s'y attarde un petit peu en introduisant les définitions et le vocabulaire spécifiques aux probabilités et aux statistiques.

2.1 Espérance d'une variable aléatoire

Définition 2.1.1 Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. L'espérance de X est

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}$$

qui est bien définie quand

- X est positif (auquel cas $\mathbb{E}[X] \in [0, +\infty]$),
- X est intégrable, i.e. $X \in L^1 = L^1(\Omega, \mathcal{F}, \mathbb{P})$.

On dit que la variable aléatoire X est *centrée* si elle est intégrable et $\mathbb{E}[X] = 0$.

Rappelons que X est intégrable si

$$\mathbb{E}[|X|] = \int_{\Omega} |X| d\mathbb{P} < +\infty,$$

ce que l'on notera en général (si le contexte est clair)

$$X \in L^1.$$

Remarque 2.1.2 Puisque \mathbb{P} est une mesure de probabilité, toute variable aléatoire bornée appartient à L^1 . En particulier, si X prend (presque sûrement) seulement un nombre fini de valeurs, alors son espérance existe et est finie.

Remarque 2.1.3 Si $\alpha \in \mathbb{R}$, alors α définit une variable aléatoire constante $\alpha \mathbf{1}_{\Omega}$. Comme cette variable aléatoire est bornée, elle appartient à L^1 . Par abus de notation, on écrira $\mathbb{E}[\alpha] := \mathbb{E}[\alpha \mathbf{1}_{\Omega}]$ et on a bien sûr $\mathbb{E}[\alpha] = \alpha$.

Remarque 2.1.4 On interprète en général l'espérance comme la valeur moyenne prise par la variable aléatoire X : lorsque Ω est fini et \mathbb{P} associe la même probabilité à tous les singletons, alors $\mathbb{E}[X]$ est bien la moyenne au sens usuel. En effet, si $\Omega = \{\omega_1, \dots, \omega_n\}$ et $\mathbb{P} = \frac{1}{n} \sum_{j=1}^n \delta_{\omega_j}$, alors

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} = \frac{1}{n} \sum_{j=1}^n X(\omega_j).$$

Les propriétés de l'espérance découlent directement des propriétés des intégrales.

Théorème 2.1.5 Soient $X, Y \in L^1$ et $\alpha \in \mathbb{R}$. Alors

- Si $B \in \mathcal{F}$, alors $\mathbb{E}[\mathbf{1}_B] = \mathbb{P}(B)$.
- *Monotonie* : Si $X \leq Y$ presque sûrement, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- *Linéarité* : $\alpha X \in L^1$ et $X + Y \in L^1$. De plus,

$$\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X] \quad \text{et} \quad \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

- *Contraction* : $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
- *Inégalité de Markov* : Si X est une variable aléatoire positive, alors

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \quad \text{pour tout } a > 0.$$

En particulier si X est positive et $\mathbb{E}[X] = 0$, alors $X = 0$ presque sûrement.

Démonstration : Cela découle de la Proposition A.5.8. Pour l'inégalité de Markov, on remarque que $\mathbf{1}_{\{X \geq a\}} \leq X/a$. Donc par monotonie $\mathbb{P}(X \geq a) = \mathbb{E}[\mathbf{1}_{\{X \geq a\}}] \leq \mathbb{E}[X/a]$. ■

Remarque 2.1.6 On peut décliner l'inégalité de Markov pour des variables aléatoires pas nécessairement positives de la manière suivante : pour tout $a, p, t > 0$

- si $|X|^p \in L^1$, alors

$$\mathbb{P}(|X| \geq a) = \mathbb{P}(|X|^p \geq a^p) \leq \frac{\mathbb{E}[|X|^p]}{a^p},$$

- si $e^{tX} \in L^1$, alors

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq e^{-ta} \mathbb{E}[e^{tX}]$$

et si $e^{-tX} \in L^1$, alors

$$\mathbb{P}(X \leq a) = \mathbb{P}(e^{-tX} \geq e^{-ta}) \leq e^{ta} \mathbb{E}[e^{-tX}].$$

Puisque l'espérance est un cas particulier d'intégrale par rapport à une mesure finie, on peut aussi lui appliquer les théorèmes de convergence rappelés dans le Chapitre A :

- Convergence monotone : si $(X_n)_{n \in \mathbb{N}}$ est une suite croissante de variables aléatoires positives qui converge (presque sûrement) vers X , alors $\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.
- Lemme de Fatou : si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires positives, alors

$$\mathbb{E}[\liminf_{n \rightarrow +\infty} X_n] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n].$$

- Convergence dominée : si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires qui converge (presque sûrement) vers X et pour laquelle il existe une variable aléatoire $Z \in L^1$ telle que $|X_n| \leq Z$ presque sûrement, alors $\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

La plupart des quantités d'intérêt associées à une variable aléatoire s'obtiennent par calcul d'une espérance du type $\mathbb{E}[h(X)]$. Pour cela, on impose que $h : \mathbb{R} \rightarrow \mathbb{R}$ soit une application mesurable (pour assurer $h(X)$ soit bien une variable aléatoire). Par définition de l'espérance de la variable aléatoire $h(X) : \Omega \rightarrow \mathbb{R}$, on a

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X) d\mathbb{P}.$$

En appliquant le Théorème A.6.3 de transfert, on peut calculer cette espérance sans connaître la loi de $h(X)$, comme le montre le résultat suivant.

Théorème 2.1.7 *Soit X une variable aléatoire et soit $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ une application mesurable.*

- Si $h : \mathbb{R} \rightarrow [0, +\infty]$, alors

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h d\mathbb{P}_X.$$

- On a $h(X) \in L^1$ si et seulement si h est \mathbb{P}_X -intégrable et dans ce cas, l'égalité précédente est vérifiée.

Le théorème de transfert nous est précieux car il permet de définir l'espérance d'une (fonction d'une) variable aléatoire sans passer par l'espace probabilisé sous-jacent, en intégrant "simplement" sur l'espace réel. Il montre en particulier que l'espérance ne dépend que de la loi de la variable aléatoire.

Corollaire 2.1.8 *Une variable aléatoire X est intégrable si et seulement si $id : \mathbb{R} \rightarrow \mathbb{R}$ est \mathbb{P}^X -intégrable et dans ce cas,*

$$\mathbb{E}[X] = \int_{\mathbb{R}} id d\mathbb{P}_X = \int_{\mathbb{R}} x d\mathbb{P}_X(x).$$

Notons les faits importants suivants. Premièrement, si X est une variable aléatoire discrète, alors si elle existe, son espérance est la moyenne pondérée de ses valeurs, où le poids est donné par la probabilité que X soit égal à la valeur considérée.

Proposition 2.1.9 (LOTUS) *Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une application mesurable et soit X une variable aléatoire discrète prenant presque sûrement ses valeurs dans un ensemble Λ . Alors, $h(X) \in L^1$ si et seulement si*

$$\sum_{x \in \Lambda} |h(x)| \mathbb{P}(X = x) < +\infty.$$

Dans ce cas, on a

$$\mathbb{E}[h(X)] = \sum_{x \in \Lambda} h(x) \mathbb{P}(X = x).$$

Démonstration : Par le Théorème 2.1.7, on sait que $h(X) \in L^1$ si et seulement si h est \mathbb{P}_X -intégrable. Or, comme X est discret, on sait que

$$\mathbb{P}_X = \sum_{x \in \Lambda} \mathbb{P}(X = x) \delta_x$$

et donc si elle existe,

$$\int_{\mathbb{R}} h \, d\mathbb{P}_X = \sum_{x \in \Lambda} h(x) \mathbb{P}(X = x).$$

La définition de l'intégrale des fonctions non-positives permet de conclure. ■

Remarque 2.1.10 Supposons que X est une variable aléatoire qui prend un nombre fini de valeurs différentes. Alors $h(X(\Omega))$ est évidemment fini. Si on écrit $X(\Omega) = \{x_1, \dots, x_n\}$ et $h(X(\Omega)) = \{y_1, y_2, \dots, y_m\}$ (avec $m \leq n$), on a $X(\Omega) = \cup_{i=1}^m h^{-1}(\{y_i\})$ et par σ -additivité, il vient

$$\mathbb{P}(h(X) = y_i) = \mathbb{P}\left(\bigcup_{j: x_j \in h^{-1}(\{y_i\})} \{X = x_j\}\right) = \sum_{j: x_j \in h^{-1}(\{y_i\})} \mathbb{P}(X = x_j)$$

et en utilisant la Proposition 2.1.9, on obtient

$$\begin{aligned} \mathbb{E}[h(X)] &= \sum_{i=1}^m y_i \mathbb{P}(h(X) = y_i) = \sum_{i=1}^m y_i \left(\sum_{j: x_j \in h^{-1}(\{y_i\})} \mathbb{P}(X = x_j) \right) \\ &= \sum_{i=1}^m \sum_{j: x_j \in h^{-1}(\{y_i\})} h(x_j) \mathbb{P}(X = x_j) = \sum_{j=1}^n h(x_j) \mathbb{P}(X = x_j), \end{aligned}$$

ce qui redémontre le résultat. Etant assez intuitif, ce résultat est souvent pris comme définition plutôt que pour conséquence du théorème de transfert, d'où son nom "Law Of The Unconscious Statistician".

Corollaire 2.1.11 Si $X \in L^1$ est une variable aléatoire discrète prenant presque sûrement ses valeurs dans un ensemble Λ , alors

$$\mathbb{E}[X] = \sum_{x \in \Lambda} x \mathbb{P}(X = x).$$

Exemple 2.1.12 Si $X \sim \text{Bern}(p)$, on a $\mathbb{P}_X = p\delta_1 + (1-p)\delta_0$ et donc

$$\mathbb{E}[X] = p \cdot 1 + (1-p) \cdot 0 = p.$$

Exemple 2.1.13 Si $X \sim \text{Unif}\{1, \dots, n\}$, alors $\mathbb{P}_X = \frac{1}{n} \sum_{j=1}^n \delta_j$ et donc

$$\mathbb{E}[X] = \sum_{j=1}^n \frac{1}{n} j = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Exemple 2.1.14 Si $X \sim \text{Poi}(\lambda)$, alors $\mathbb{P}_X = \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$ et

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{\lambda^k}{k!} k = e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Le deuxième fait important est le calcul de l'espérance de variables aléatoires continues.

Proposition 2.1.15 (LOTUS) Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une application mesurable et soit X une variable aléatoire continue de densité f . Alors, $h(X) \in L^1$ si et seulement si hf est intégrable sur \mathbb{R} . Dans ce cas,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx.$$

Démonstration : Le Théorème 2.1.7 implique que

$$\mathbb{E}[h(X)] = \int_{\Omega} h \circ X d\mathbb{P} = \int_{\mathbb{R}} h d\mathbb{P}_X.$$

On applique ensuite la Proposition A.7.3 pour écrire

$$\mathbb{E}[h(X)] = \int h d\mathbb{P}_X = \int h d(f \cdot \lambda) = \int hf d\lambda = \int_{\mathbb{R}} h(x)f(x)dx.$$

■

Corollaire 2.1.16 Supposons que X est une variable aléatoire continue de densité f . Alors, $X \in L^1$ si et seulement si la fonction $x \mapsto xf(x)$ est intégrable sur \mathbb{R} . Dans ce cas, on a

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx.$$

Exemple 2.1.17 Si $X \sim \text{Unif}(a, b)$, alors $f = \frac{1}{b-a} \mathbb{1}_{[a,b]}$ et donc

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

Exemple 2.1.18 Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ et donc

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sigma u + \mu) e^{-\frac{u^2}{2}} du \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sigma u e^{-\frac{u^2}{2}} du + \frac{\mu}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}} du = \mu \end{aligned}$$

en effectuant le changement de variable $u = \frac{x-\mu}{\sigma}$ et en utilisant l'intégrale de Poisson.

Exemple 2.1.19 (LOTUS mixte) Considérons un exemple de loi mixte ni discrète ni continue, en prenant $N \sim \mathcal{N}(0, 1)$ et

$$X = N \mathbb{1}_{\{N \leq 1\}} + \mathbb{1}_{\{N > 1\}}$$

qui correspond à une gaussienne tronquée en 1. Ce qui rend cette loi mixte est le fait qu'elle prend ses valeurs à la fois dans un intervalle *et* elle a un point de probabilité non nulle. Plus précisément, si nous notons Φ la fonction de répartition gaussienne standard, alors

$$F(x) = \Phi(x)\mathbb{1}_{]-\infty,1[} + \mathbb{1}_{[1,+\infty[}.$$

Cette fonction est discontinue au point $x = 1$, constante sur $]1, +\infty[$ et strictement croissante ailleurs. Il y a un point de discontinuité de la fonction de répartition au point $x = 1$, avec un saut $F(1) - F(1^-) = 1 - \Phi(1)$ ce qui indique que $\mathbb{P}(X = 1) = 1 - \Phi(1)$ (qui n'est donc pas nulle). En notant ϕ la densité gaussienne standard, on calcule

$$\mathbb{E}[h(X)] = \int_{-\infty}^1 h(x)\phi(x)dx + h(1)(1 - \Phi(1)).$$

De façon générale, les espérances de lois mixtes se calculent de la même façon, grosso modo en intégrant partout où la cdf est continue et en sommant là où la loi a des sauts.

Terminons cette section par une généralisation du LOTUS au cas de vecteurs aléatoires. En effet, si X est un vecteur aléatoire et si $h : \mathbb{R}^d \rightarrow \mathbb{R}$ est mesurable, alors $h(X)$ définit une variable aléatoire. On peut s'intéresser au calcul de son espérance. On retrouve à nouveau la formule de calcul suivante.

Théorème 2.1.20 (LOTUS) Soit (X_1, \dots, X_d) un vecteur aléatoire et $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ une application mesurable.

- Si $h : \mathbb{R}^d \rightarrow [0; +\infty]$, alors

$$\mathbb{E}[h(X_1, \dots, X_d)] = \int_{\mathbb{R}^d} h d\mathbb{P}_{(X_1, \dots, X_d)}.$$

- On a $h(X_1, \dots, X_d) \in L^1$ si et seulement si h est $\mathbb{P}_{(X_1, \dots, X_d)}$ -intégrable et dans ce cas, l'égalité précédente est vérifiée.

En particulier,

- si le vecteur aléatoire (X_1, \dots, X_d) est discret et si pour tout $j \in \{1, \dots, d\}$, X_j prend presque sûrement ses valeurs dans un ensemble Λ_j , alors

$$\mathbb{E}[h(X_1, \dots, X_d)] = \sum_{j_1 \in \Lambda_1} \dots \sum_{j_d \in \Lambda_d} h(x_{j_1}, \dots, x_{j_d}) \mathbb{P}(X_1 = x_{j_1}, \dots, X_d = x_{j_d});$$

- si le vecteur aléatoire (X_1, \dots, X_d) est continu et admet la fonction de densité f , alors

$$\mathbb{E}[h(X_1, \dots, X_d)] = \int_{\mathbb{R}^d} h(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \dots dx_d.$$

2.2 Moments et variance

Un exemple important d'application $h : \mathbb{R} \rightarrow \mathbb{R}$ est la fonction $h(x) = x^p$. Cela nous amène à considérer les *moments* d'une variable aléatoire et les espaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Notons que puisque la mesure \mathbb{P} est finie, on a

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^1(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{pour tout } p \geq 1.$$

En effet, si $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, alors les inégalités de Hölder donnent

$$\mathbb{E}[|X|] = \mathbb{E}[|X|1] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[1^q]^{\frac{1}{q}} = \mathbb{E}[|X|^p]^{\frac{1}{p}}$$

si p et q sont des exposants conjugués. De même, on trouve que

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^r(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{pour tous } p \geq r \geq 1.$$

Dans le cas particulier où $p = 2$, on trouve la formule très utile

$$\mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2].$$

De manière générale, on s'intéresse aux moments de la variable aléatoire considérée (souvent pour les valeurs de p entières).

Définition 2.2.1 Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $p \in \mathbb{N}_0$. On dit que X admet un moment d'ordre p si $X \in L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$. Dans ce cas, le moment d'ordre p de X est défini par

$$\mathbb{E}[X^p] = \int_{\Omega} X^p d\mathbb{P}.$$

En particulier, le moment d'ordre 1 est l'espérance de X .

Définition 2.2.2 Soient $p \geq 1$ et $X \in L^p$. Le moment centré de X d'ordre p de X est défini par

$$\mu_p = \mathbb{E}[(X - \mathbb{E}[X])^p].$$

Remarquons que si $X \in L^p$, alors $X - \mathbb{E}[X] \in L^p$ puisque la mesure de probabilité est finie. Le moment centré d'ordre p est donc bien défini.

Remarque 2.2.3 Si $X \in L^1$ alors $\mu_1 = 0$. A ce titre, on appelle l'opération $X \mapsto X - \mathbb{E}[X]$ un centrage de la variable aléatoire.

Un cas particulier important est $p = 2$.

Définition 2.2.4 Soit $X \in L^2$. Le moment centré d'ordre 2 de X est appelé la variance de X et est noté $\text{Var}[X]$. L'écart-type est $\sigma_X = \sqrt{\text{Var}[X]}$.

Attention, la notation σ_X n'a rien à voir avec la σ -algèbre engendrée par X . Les propriétés de la variance sont rassemblées dans les résultats suivants.

Proposition 2.2.5 (Propriétés de la variance) Soit $X \in L^2$ et soit $a \in \mathbb{R}$. Alors

- $\text{Var}[X] \geq 0$,
- $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ et, en particulier, $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$,
- $\text{Var}[X + a] = \text{Var}[X]$,
- $\text{Var}[aX] = a^2 \text{Var}[X]$,
- $\text{Var}[a] = 0$,
- $\text{Var}[X] = 0$ si et seulement si $X = \mathbb{E}[X]$ presque sûrement.

Démonstration : Tout est évident. Par exemple, par linéarité de l'espérance, on a

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X\mathbb{E}[X]] + (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

■

Exemple 2.2.6 Si $X \sim \text{Bern}(p)$, on sait que $\mathbb{E}[X] = p$. On a aussi que $X = X^2$ et donc $\mathbb{E}[X^2] = p$. Il s'ensuit que

$$\text{Var}[X] = p - p^2 = p(1 - p).$$

Exemple 2.2.7 Si $X \sim \text{Unif}\{1, \dots, n\}$, on a $\mathbb{E}[X] = \frac{n+1}{2}$. On calcule alors

$$\mathbb{E}[X^2] = \sum_{j=1}^n \frac{1}{n} j^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{3} = \frac{(n+1)(2n+1)}{3}.$$

Ainsi,

$$\text{Var}[X] = \frac{(n+1)(2n+1)}{3} - \frac{(n+1)^2}{4} = \frac{n^2 - 1}{12}.$$

Exemple 2.2.8 Si $X \sim \text{Poi}(\lambda)$, alors on sait que $\mathbb{E}[X] = \lambda$. De même, on calcule

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{\lambda^k}{k!} k^2 = \lambda \sum_{k=1}^{+\infty} e^{-\lambda} k \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda \sum_{k=0}^{+\infty} e^{-\lambda} (k+1) \frac{\lambda^k}{k!} \\ &= \lambda \left(\sum_{k=0}^{+\infty} e^{-\lambda} k \frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} \right) \\ &= \lambda(\mathbb{E}[X] + 1) = \lambda(\lambda + 1) \end{aligned}$$

d'où

$$\text{Var}[X] = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

Exemple 2.2.9 Si $X \sim \text{Unif}(a, b)$, alors on sait que $\mathbb{E}[X] = \frac{a+b}{2}$. De plus, on calcule

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

d'où

$$\text{Var}[X] = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Exemple 2.2.10 Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors on a $\mathbb{E}[X] = \mu$. Montrons que $\text{Var}[X] = \sigma^2$. On a

$$\begin{aligned} \text{Var}[X] = \mathbb{E}[(X - \mu)^2] &= \int_{\mathbb{R}} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} u^2 e^{-\frac{u^2}{2}} du \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left([-ue^{-u^2/2}]_{-\infty}^{+\infty} + \int_{\mathbb{R}} e^{-\frac{u^2}{2}} du \right) \\ &= \sigma^2 \end{aligned}$$

en effectuant le changement de variable $u = \frac{x-\mu}{\sigma}$, une intégration par parties et en utilisant l'intégrale de Poisson.

L'espérance $\mathbb{E}[X]$ d'une variable aléatoire X nous renseigne sur la valeur "moyenne" ou "attendue" de X . Ceci se voit très clairement dans le cas d'une variable aléatoire discrète comme déjà signalé. On le voit de façon générale grâce au théorème suivant.

Théorème 2.2.11 (Décomposition de König-Huygens) Si $X \in L^2$, alors

$$\mathbb{E}[(X - a)^2] = \text{Var}[X] + (\mathbb{E}[X] - a)^2 \text{ pour tout } a \in \mathbb{R}.$$

En particulier,

$$\mathbb{E}[X] = \operatorname{argmin}_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2]$$

c'est-à-dire l'espérance de X est la meilleure approximation de X en norme L^2 par une constante. De plus, sa variance est l'écart minimal qu'on peut obtenir.

Démonstration : On utilise les propriétés de l'espérance pour écrire

$$\begin{aligned} \mathbb{E}[(X - a)^2] &= \mathbb{E} \left[((X - \mathbb{E}[X]) + (\mathbb{E}[X] - a))^2 \right] \\ &= \mathbb{E} \left[(X - \mathbb{E}[X])^2 \right] + \mathbb{E} [2(X - \mathbb{E}[X])(\mathbb{E}[X] - a)] + \mathbb{E}[(\mathbb{E}[X] - a)^2] \\ &= \mathbb{E} \left[(X - \mathbb{E}[X])^2 \right] + 0 + (\mathbb{E}[X] - a)^2. \end{aligned}$$

La conclusion suit. ■

Le résultat suivant montre que la variance est une mesure de la dispersion de X autour de son espérance.

Théorème 2.2.12 (Inégalité de Bienaymé-Tchebychev) Si $X \in L^2$ et $a > 0$, alors

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}. \tag{2.1}$$

Démonstration : Cela découle directement de l'inégalité de Markov appliquée à la variable aléatoire positive $(X - \mathbb{E}[X])^2$. ■

Ainsi, seule une grande variance peut permettre à une variable X de différer significativement de son espérance puisque dans ce cas, le risque de voir une valeur extrême de X augmente. On peut

interpréter cette inégalité de la manière suivante : la variable aléatoire X prend ses valeurs dans l'intervalle $[\mathbb{E}[X] - a\sigma_X, \mathbb{E}[X] + a\sigma_X]$ avec une probabilité plus grande ou égale à $(1 - \frac{1}{a^2})$. Pour $a = 2$, on trouve une probabilité d'au moins $\frac{3}{4}$ d'être dans l'intervalle $[\mathbb{E}[X] - 2\sigma_X, \mathbb{E}[X] + 2\sigma_X]$. Pour $a = 3$, la probabilité d'être dans l'intervalle $[\mathbb{E}[X] - 3\sigma_X, \mathbb{E}[X] + 3\sigma_X]$ est supérieure ou égale à $8/9 \approx 0,888$.

2.3 Covariance et corrélation

Si X et Y sont deux variables aléatoires de L^1 , on sait par linéarité de l'espérance que

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Si X et Y appartiennent à L^2 , on sait que $X + Y \in L^2$ (puisque L^2 est un espace vectoriel, cela découle de l'inégalité de Minkowski). Par contre, en général, on a

$$\text{Var}[X + Y] \neq \text{Var}[X] + \text{Var}[Y].$$

Il suffit de prendre $X = Y$ puisque $\text{Var}[2X] = 4\text{Var}[X]$. Un second exemple est donné ci-dessous.

Exemple 2.3.1 Supposons que $X \sim \text{Bern}(p)$ et posons $Y = 1 - X$. Alors $X + Y$ est la variable aléatoire constante égale à 1 et donc $\text{Var}[X + Y] = 0$. Or, on sait que $\text{Var}[X] = p(1 - p)$ et $\text{Var}[Y] = p(1 - p)$, d'où $\text{Var}[X] + \text{Var}[Y] \neq \text{Var}[X + Y]$.

En fait, les propriétés de la variance montrent que l'application Var est une *forme quadratique* sur L^2 ; la forme bilinéaire associée est donnée par

$$(X, Y) \mapsto \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Définition 2.3.2 Si $X, Y \in L^2$, la *covariance* de X et Y est définie par

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Remarque 2.3.3 L'interprétation de la covariance est illustrée dans les Figures 2.1 et 2.2.

En pratique, on utilise le résultat suivant pour calculer des covariances : celui-ci montre en particulier que l'espérance d'un produit n'est en général pas égal au produit des espérances.

Proposition 2.3.4 Si $X, Y \in L^2$, alors

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Démonstration : Il suffit de développer le produit $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$ et d'utiliser la linéarité de l'espérance. ■

La covariance est une forme bilinéaire symétrique semi-définie positive sur L^2 . La formule de polarisation des formes bilinéaires (qui généralise l'égalité $2ab = (a + b)^2 - a^2 - b^2$) donne directement le résultat suivant.

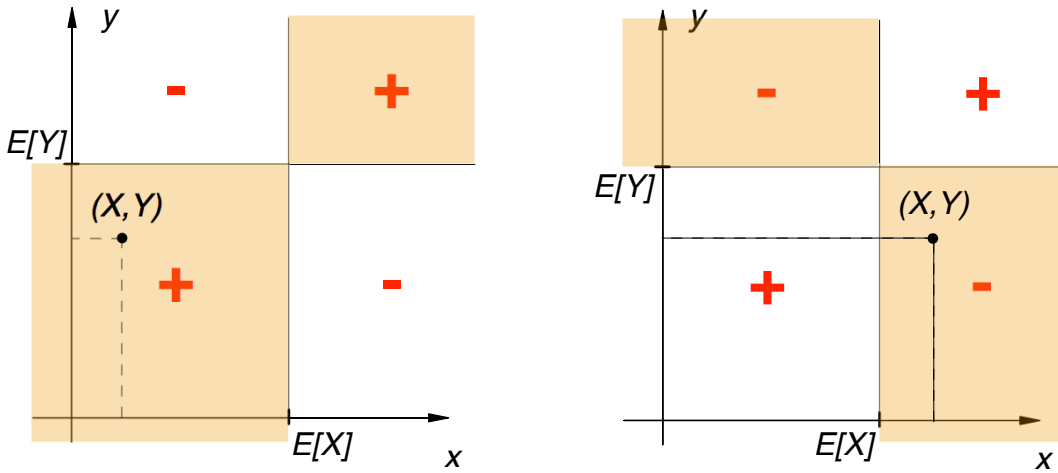


FIGURE 2.1 – A gauche, (X, Y) se réalise dans la zone où $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$. A droite, (X, Y) se réalise dans la zone où $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$. Si (X, Y) se réalise plus souvent dans les zones "+" (ou de façon plus extrême), $\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ sera positif, et inversement.

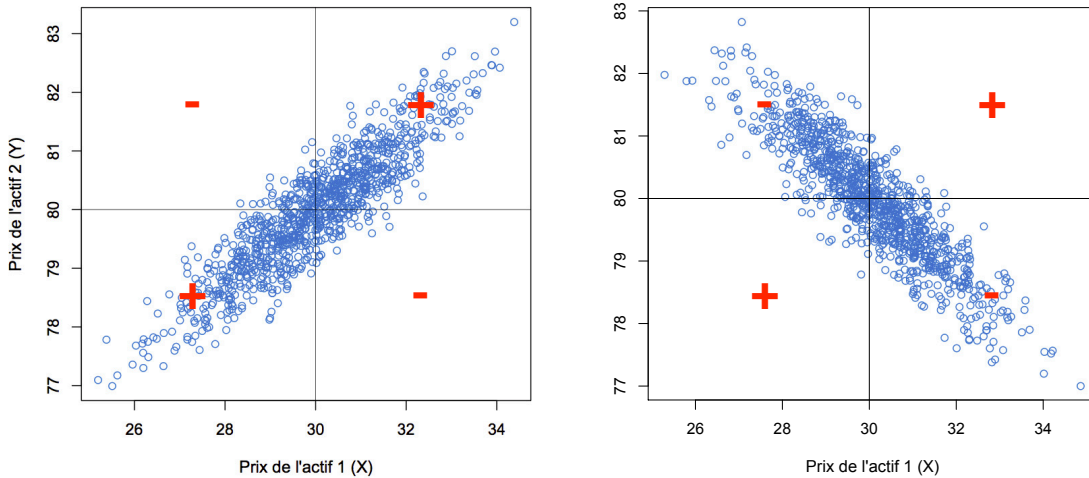


FIGURE 2.2 – A gauche, $\text{Cov}[X, Y] > 0$; on parlera de dépendance positive (dans ce cas, on a $\text{Var}[X + Y] > \text{Var}[X] + \text{Var}[Y]$). A droite, $\text{Cov}[X, Y] < 0$; on parlera de dépendance négative (dans ce cas, $\text{Var}[X + Y] < \text{Var}[X] + \text{Var}[Y]$).

Proposition 2.3.5 Si $X, Y \in L^2$, alors

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y].$$

Démonstration : La formule de polarisation donne

$$\text{Cov}[X, Y] = \frac{1}{2} (\text{Var}[X + Y] - \text{Var}[X] - \text{Var}[Y])$$

d'où le résultat. ■

Les résultats suivants proviennent directement de la bilinéarité de la covariance.

Proposition 2.3.6 Soient $X, Y, Z \in L^2$ et $c \in \mathbb{R}$. Alors

- $\text{Cov}[X, X] = \text{Var}[X]$
- $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
- $\text{Cov}[cX, Y] = c \text{Cov}[X, Y]$
- $\text{Cov}[X + Y, Z] = \text{Cov}[X, Z] + \text{Cov}[Y, Z]$
- $\text{Cov}[X, c] = 0$
- $\text{Cov}[X + c, Y] = \text{Cov}[X, Y]$.

Si on travaille sur le sous-espace de L^2 formé des variables aléatoires centrées quotienté par la relation d'équivalence donnée par l'égalité presque sûre, alors la covariance y est une forme bilinéaire définie positive. On retrouve donc l'inégalité de Cauchy-Schwarz

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

si $X, Y \in L^2$ sont tels que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. De plus, on a l'égalité si et seulement si X et Y sont linéairement dépendants. Notons qu'on peut toujours se ramener à des variables aléatoires centrées en remplaçant X par $X - \mathbb{E}[X]$. Nous en rappelons néanmoins la preuve.

Proposition 2.3.7 (Inégalité de Cauchy-Schwarz) Soient $X, Y \in L^2$. Alors

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]\text{Var}[Y]}$$

avec égalité si et seulement si $Y = \alpha X + \beta$ ou $X = \alpha Y + \beta$ presque sûrement pour certains $\alpha, \beta \in \mathbb{R}$ et $\text{sign}(\alpha) = \text{sign}(\text{Cov}[X, Y])$.

Démonstration : Si X, Y sont centrées, alors

$$0 \leq \mathbb{E}[(X + \lambda Y)^2] = \mathbb{E}[X^2] + 2\lambda \mathbb{E}[XY] + \lambda^2 \mathbb{E}[Y^2]$$

pour tout $\lambda \in \mathbb{R}$. On a donc un polynôme de degré 2 en λ qui est toujours positif ou nul et donc

$$4\mathbb{E}[XY]^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$$

c'est-à-dire

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

On a l'égalité si et seulement si le polynôme admet une racine, c'est-à-dire il existe $\lambda \in \mathbb{R}$ tel que $\mathbb{E}[(X + \lambda Y)^2] = 0$. On en tire que $X + \lambda Y = 0$ presque sûrement. On obtient la conclusion générale appliquant le résultat aux versions centrées de X et Y . ■

Remarque 2.3.8 On montre directement que si $Y = \alpha X + \beta$, on a

$$\alpha = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} \quad \text{et} \quad \beta = \mathbb{E}[Y] - \alpha \mathbb{E}[X].$$

Ceci amène naturellement à introduire la notion suivante.

Définition 2.3.9 Soient $X, Y \in L^2$. La *corrélation* entre X et Y est la quantité

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}.$$

On prend la convention que la corrélation est nulle si $\text{Var}[X] = 0$ ou $\text{Var}[Y] = 0$, ce qui revient à dire qu'une des deux variables aléatoires est presque-sûrement constante. Vu l'inégalité de Cauchy-Schwarz, la corrélation rend compte de la dépendance *linéaire*.

Définition 2.3.10 Soient $X, Y \in L^2$. Si $\text{Corr}[X, Y] = 0$, on dit que X et Y sont *non-corrélées*.

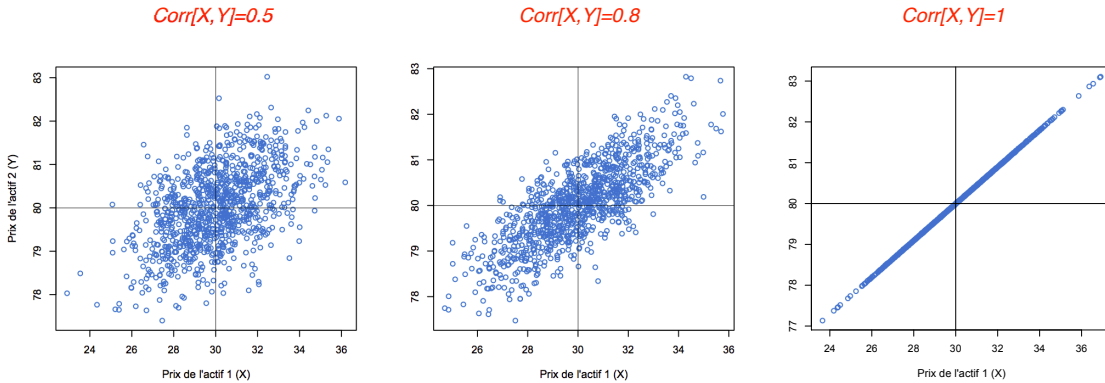


FIGURE 2.3 – Plus $|\text{Corr}[X, Y]|$ est proche de 1, plus la relation linéaire entre X et Y est forte.

Il est à noter que les signes de $\text{Corr}[X, Y]$ et de $\text{Cov}[X, Y]$ coïncident. La corrélation satisfait les propriétés suivantes.

Proposition 2.3.11 Soient X, Y, X_1, X_2, Y_1, Y_2 des variables aléatoires et $c_1, c_2, d_1, d_2 \in \mathbb{R}$. Alors

- $\text{Corr}[c_1X + d_1, c_2Y + d_2] = \text{Corr}[X, Y]$
- $\text{Corr}[X, Y] = \text{Corr}[Y, X]$
- $\text{Corr}[X, c] = 0$
- $\text{Corr}[X, X] = 1$
- $|\text{Corr}[X, Y]| \leq 1$, et l'égalité a lieu si et seulement si $Y = \alpha X + \beta$ p.s. ou $X = \alpha Y + \beta$ p.s. pour certains $\alpha \in \mathbb{R}_0, \beta \in \mathbb{R}$.

2.4 Moments de vecteurs aléatoires

On généralise la définition de l'espérance au cas des vecteurs aléatoires comme suit. On prend la convention que nos vecteurs sont des vecteurs colonnes.

Définition 2.4.1 Si $X = (X_1, \dots, X_d)^T$ est un vecteur aléatoire, on définit l'*espérance* ou le *vecteur moyen* de X par

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^T$$

si chacune de ces espérances est bien définie. On dit que X est *intégrable*, ce que l'on note $X \in L^1$, si $X_j \in L^1$ pour tout $j \in \{1, \dots, d\}$.

Comme l'espérance d'un vecteur aléatoire est donnée par les espérances de chacune de ses composantes, les propriétés du vecteur moyen découlent de celles de l'espérance d'une variable aléatoire. De même, on peut définir l'espérance d'une matrice formée de variables aléatoires en prenant l'espérance composante à composante.

Intéressons-nous maintenant aux propriétés de covariance entre les variables aléatoires d'un vecteur aléatoire colonne $X = (X_1, \dots, X_d)^T$.

Définition 2.4.2 Si $X = (X_1, \dots, X_d)^T$ est un vecteur aléatoire de L^2 (i.e. chaque composante appartient à L^2), on définit la *variance* de X (aussi appelée la *matrice de variance-covariance*) par

$$\text{Var}[X] = \mathbb{E} [(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T].$$

La variance de X est donc une matrice carrée $d \times d$. En développant le produit matriciel, il vient

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] \\ &= \mathbb{E} \left[\begin{pmatrix} X_1 - \mathbb{E}[X_1] \\ X_2 - \mathbb{E}[X_2] \\ \vdots \\ X_d - \mathbb{E}[X_d] \end{pmatrix} \begin{pmatrix} X_1 - \mathbb{E}[X_1], X_2 - \mathbb{E}[X_2], \dots, X_d - \mathbb{E}[X_d] \end{pmatrix} \right] \\ &= \begin{pmatrix} \text{Var}[X_1] & \dots & \text{Cov}[X_1, X_d] \\ \vdots & \ddots & \vdots \\ \text{Cov}[X_d, X_1] & \dots & \text{Var}[X_d] \end{pmatrix} \end{aligned}$$

c'est-à-dire

$$\text{Var}[X]_{ij} = \text{Cov}[X_i, X_j].$$

Cela justifie le nom donné à cette matrice. Le résultat suivant s'obtient également par simple manipulation matricielle.

Les propriétés de l'espérance et du produit matriciel nous amènent les résultats utiles suivants.

Proposition 2.4.3 Soit $X \in L^2$ un vecteur aléatoire de \mathbb{R}^d . Soient A, B des matrices de dimensions compatibles. Alors

$$\mathbb{E}[AX + B] = A\mathbb{E}[X] + B \text{ et } \text{Var}[AX + B] = A\text{Var}[X]A^T.$$

En particulier, pour tout $b \in \mathbb{R}^d$, on a

$$b^T \text{Var}[X] b = \text{Var}[b^T X] \geq 0,$$

donc une matrice de variance-covariance est toujours semi-définie positive.

Remarque 2.4.4 Soit X un vecteur aléatoire de dimension d . Si elle existe, la matrice $\text{Var}[X]$ est symétrique et semi-définie positive. On peut donc la représenter sous la forme $\text{Var}[X] = AA^T$ avec A une matrice carrée (on dit que A est une racine carrée de $\text{Var}[X]$)¹.

- *Centrer* la loi de X revient à considérer le vecteur $Y = X - \mathbb{E}[X]$ qui, par construction, est d'espérance $\mathbb{E}[Y] = 0$. Cette opération est permise dès que X est intégrable.
- *Réduire* la loi de X revient à considérer le vecteur $Z = A^{-1}X$ qui, par construction, est de variance $\text{Var}[Z] = \text{Id}$. Cette opération est permise dès que X est de carré intégrable et que $\text{Var}[X]$ est définie positive.

2.5 Formules et inégalités liées à l'espérance

Nous avons déjà démontré l'inégalité de Markov et ses conséquences. L'objectif de cette section est de démontrer quelques autres (in)égalités utiles.

Proposition 2.5.1 Soit X une variable aléatoire positive et soit $h : [0, +\infty[\rightarrow [0, +\infty[$ une fonction croissante continûment dérivable. Alors

$$\mathbb{E}[h(X)] = h(0) + \int_0^{+\infty} \mathbb{P}(X > t) h'(t) dt = h(0) + \int_0^{+\infty} \mathbb{P}(X \geq t) h'(t) dt.$$

Démonstration : Comme la variable aléatoire $h(X)$ est positive, son espérance existe. De plus, on a

$$h(x) = h(0) + \int_0^x h'(t) dt$$

pour tout $x \geq 0$ puisque h est continûment dérivable. Il s'ensuit que

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_{\Omega} h(X) d\mathbb{P} = \int_{\Omega} \left(h(0) + \int_0^{X(\omega)} h'(t) dt \right) d\mathbb{P}(\omega) \\ &= h(0) + \int_{\Omega} \left(\int_0^{+\infty} \mathbb{1}_{\{X > t\}}(\omega) h'(t) dt \right) d\mathbb{P}(\omega) \\ &= h(0) + \int_0^{+\infty} h'(t) \left(\int_{\Omega} \mathbb{1}_{\{X > t\}}(\omega) d\mathbb{P}(\omega) \right) dt \\ &= h(0) + \int_0^{+\infty} h'(t) \mathbb{P}(X > t) dt \end{aligned}$$

puisque $\mathbb{P}(\Omega) = 1$ et en utilisant le théorème de Fubini (pour les fonctions positives puisque h est croissant). La preuve de la deuxième égalité est identique puisque $\mathbb{P}(X > t) = \mathbb{P}(X \geq t)$ pour λ -presque tout $t \geq 0$ par la Proposition 1.3.13. ■

1. Comme $\text{Var}[X]$ est symétrique, on peut la diagonaliser par une matrice orthogonale P . Ainsi, on peut écrire $\text{Var}[X] = P^T \text{diag}(\lambda_1, \dots, \lambda_d) P$. De plus, comme la matrice $\text{Var}[X]$ est semi-définie positive, ses valeurs propres $\lambda_1, \dots, \lambda_d$ sont positives ou nulles. Il suffit alors de poser $A = P^T \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$.

Corollaire 2.5.2 Soit X une variable aléatoire. Pour tout $p \geq 1$, on a

$$\mathbb{E}[|X|^p] = p \int_0^{+\infty} t^{p-1} \mathbb{P}(|X| > t) dt = p \int_0^{+\infty} t^{p-1} \mathbb{P}(|X| \geq t) dt.$$

En particulier,

$$\mathbb{E}[|X|] = \int_0^{+\infty} \mathbb{P}(|X| > t) dt = \int_0^{+\infty} \mathbb{P}(|X| \geq t) dt.$$

Démonstration : Il suffit d'appliquer la Proposition 2.5.1 avec la fonction $h(x) = x^p$ et la variable aléatoire $|X|$. ■

Ce résultat permet d'exprimer l'intégrabilité d'une variable aléatoire positive X à partir de sa fonction de répartition puisque

$$\mathbb{E}[X] = \int_0^{+\infty} 1 - F_X(t) dt.$$

Ainsi $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si $1 - F_{|X|} \in L^1([0, +\infty[)$. On peut également ré-exprimer cette condition sous la forme suivante :

$$X \in L^1 \iff \sum_{n=0}^{+\infty} \mathbb{P}(|X| > n) < +\infty \quad (2.2)$$

puisque

$$\mathbb{E}[|X|] = \sum_{n=0}^{+\infty} \int_n^{n+1} \mathbb{P}(|X| > t) dt \geq \sum_{n=0}^{+\infty} \mathbb{P}(|X| > n+1)$$

et

$$\mathbb{E}[|X|] = \sum_{n=0}^{+\infty} \int_n^{n+1} \mathbb{P}(|X| > t) dt \leq \sum_{n=0}^{+\infty} \mathbb{P}(|X| > n).$$

Dans le cas d'une variable aléatoires à valeurs dans \mathbb{N} , cette série nous donne exactement l'espérance de X .

Corollaire 2.5.3 Soit $n_0 \in \mathbb{N}$. Si X est une variable aléatoire à valeurs dans l'ensemble $\{n \in \mathbb{N} : n \geq n_0\}$, alors

$$\mathbb{E}[X] = n_0 + \sum_{n=n_0+1}^{+\infty} \mathbb{P}(X \geq n).$$

Démonstration : Comme X est positive, on peut calculer son espérance. Par le Corollaire 2.5.2, on a

$$\begin{aligned} \mathbb{E}[X] &= \int_0^{+\infty} \mathbb{P}(|X| > t) dt = \int_0^{n_0} 1 dt + \sum_{j=n_0+1}^{+\infty} \int_{j-1}^j \mathbb{P}(|X| > t) dt \\ &= n_0 + \sum_{j=n_0+1}^{+\infty} \mathbb{P}(|X| \geq j) \end{aligned}$$

puisque $\{X > t\} = \{X \geq j\}$ pour tout $t \in]j, j+1[$. ■

Passons à présent à l'inégalité de Jensen. Rappelons que si I est un intervalle, une fonction $\Phi : I \rightarrow \mathbb{R}$ est *convexe* sur I si

$$\Phi((1 - \lambda)x + \lambda y) \leq (1 - \lambda)\Phi(x) + \lambda\Phi(y)$$

pour tous $x, y \in I$ et tout $\lambda \in [0, 1]$.

En particulier, si Φ est une fonction convexe sur I et si X est une variable aléatoire qui prend les valeurs $x, y \in I$ avec une probabilité λ et $1 - \lambda$ respectivement, la convexité de Φ permet d'écrire $\Phi(\mathbb{E}[X]) \leq \mathbb{E}[\Phi(X)]$. L'inégalité de Jensen que nous démontrons ci-dessous permet de généraliser ce résultat à une variable aléatoire de L^1 quelconque.

Rappelons que toute fonction convexe sur un intervalle y est continue. En particulier, elle est mesurable. L'inégalité des pentes permet également de montrer que si Φ est convexe, elle admet des dérivées à gauche et à droite en chaque point. En particulier, pour tout $x \in I^\circ$, on a

$$\Phi(y) \geq \Phi(x) + \beta(y - x), \quad \forall y \in I$$

en prenant β égal à la dérivée à gauche ou à droite de Φ en x (ou n'importe quelle valeur entre les deux).

Proposition 2.5.4 (Inégalité de Jensen) *Soit $X \in L^1$ à valeurs dans un intervalle I de \mathbb{R} et soit $\Phi : I \rightarrow \mathbb{R}$ une fonction convexe telle que $\Phi(X) \geq 0$ ou $\Phi(X) \in L^1$. Alors, on a*

$$\Phi(\mathbb{E}[X]) \leq \mathbb{E}[\Phi(X)].$$

Démonstration : Notons $I = [a, b]$ et commençons par remarquer que par monotonie de l'espérance, on doit avoir $\mathbb{E}[X] \in [a, b]$. De plus, comme $X - a \geq 0$, on a $\mathbb{E}[X] = a$ si et seulement si $X = a$ presque sûrement. De même, $\mathbb{E}[X] = b$ si et seulement si $X = b$ presque sûrement. Dans le cas où la variable aléatoire X est presque sûrement constante, l'inégalité de Jensen devient alors une égalité triviale. Supposons donc que X n'est pas presque sûrement constant et donc que $\mathbb{E}[X] \in I^\circ$. Par conséquent, il existe un nombre réel β tel que

$$\Phi(y) \geq \Phi(\mathbb{E}[X]) + \beta(y - \mathbb{E}[X])$$

pour tout $y \in I$. En particulier, on a

$$\Phi(X) \geq \Phi(\mathbb{E}[X]) + \beta(X - \mathbb{E}[X])$$

et par monotonie de l'espérance, on obtient

$$\mathbb{E}[\Phi(X)] \geq \mathbb{E}[\Phi(\mathbb{E}[X]) + \beta(X - \mathbb{E}[X])] = \Phi(\mathbb{E}[X]) + \beta\mathbb{E}[X - \mathbb{E}[X]].$$

On conclut en remarquant que $\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$. ■

Remarque 2.5.5 On retrouve en particulier que $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ pour toute variable aléatoire $X \in L^1$.

Remarque 2.5.6 On peut également retrouver l'inégalité de Jensen discrète : si Φ est une fonction convexe sur un intervalle I de \mathbb{R} , alors pour tout $n \in \mathbb{N}_0$, tous $x_1, \dots, x_n \in I$ et tous $\lambda_1, \dots, \lambda_n \in [0, 1]$ tels que $\sum_{j=1}^n \lambda_j = 1$, on a

$$\Phi \left(\sum_{j=1}^n \lambda_j x_j \right) \leq \sum_{j=1}^n \lambda_j \Phi(x_j).$$

Il suffit pour cela d'appliquer l'inégalité de Jensen à une variable aléatoire X prenant les valeurs x_1, \dots, x_n avec probabilité $\lambda_1, \dots, \lambda_n$ respectivement.

Corollaire 2.5.7 Si $p \geq r \geq 1$ et si X est une variable aléatoire, alors $\mathbb{E}[|X|^r]^{\frac{1}{r}} \leq \mathbb{E}[|X|^p]^{\frac{1}{p}}$. En particulier, on a

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^r(\Omega, \mathcal{F}, \mathbb{P}).$$

Démonstration : Puisque $p \geq r$, il est clair que la fonction Φ définie sur \mathbb{R} par $\Phi(x) = x^{p/r}$ est convexe sur $[0, +\infty[$. Par conséquent, l'inégalité de Jensen appliquée à la variable aléatoire $|X|^r$ donne

$$(\mathbb{E}[|X|^r])^{p/r} \leq \mathbb{E}[|X|^p]$$

ce qui suffit. ■

2.6 Fonction caractéristique

Comme nous l'avons vu, étant donnée une variable aléatoire X , on peut calculer $\mathbb{E}[h(X)]$ pour n'importe quelle fonction mesurable h telle que $h(X) \in L^1$, et certaines fonctions apportent plus d'information que d'autres. Remarquons également que si $h = \mathbb{1}_{]-\infty, t]}$ pour $t \in \mathbb{R}$ alors $\mathbb{E}[h(X)] = \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X)] = \mathbb{P}(X \in]-\infty, t]) = F_X(t)$. Il existe donc des classes de fonctions \mathcal{G} telles que

$$X \stackrel{\mathcal{L}}{=} Y \text{ si et seulement si } \mathbb{E}[h(X)] = \mathbb{E}[h(Y)] \text{ pour tout } h \in \mathcal{G}.$$

Il est parfois plus simple d'étudier la loi par l'intermédiaire de son espérance sur une classe de fonctions.

Définition 2.6.1 Soit X un vecteur aléatoire dans \mathbb{R}^d . Sa *fonction caractéristique* est la fonction φ^X définie par

$$\varphi^X : \mathbb{R}^d \rightarrow \mathbb{C} : t \mapsto \mathbb{E} \left[e^{it^T X} \right].$$

Remarquons que puisque $|e^{it^T X}| = 1$, la variable aléatoire $e^{it^T X}$ est bien dans L^1 pour tout $t \in \mathbb{R}^d$ et l'espérance apparaissant dans la définition de φ^X existe bien.

Proposition 2.6.2 Si X est un vecteur aléatoire dans \mathbb{R}^d , alors

- sa fonction caractéristique φ^X est continue sur \mathbb{R}^d ,
- $\varphi^X(0) = 1$,
- $|\varphi^X(t)| \leq 1$ pour tout $t \in \mathbb{R}^d$.

Démonstration : Pour la continuité, on sait que si $t_n \rightarrow t$ alors $e^{it_n^T x} \rightarrow e^{it^T x}$ et donc, grâce au théorème de la convergence dominée, $\varphi^X(t_n) \rightarrow \varphi^X(t)$. La deuxième affirmation est évidente, tandis que la dernière découle immédiatement de l'inégalité de Jensen. ■

Exemple 2.6.3 Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ une v.a.r. de loi normale de moyenne μ et de variance σ^2 . Alors $\varphi^X(t) = \exp(it\mu - \sigma^2 t^2/2)$.

En écrivant $\varphi^X(t) = \int_{\mathbb{R}^d} e^{it^T x} d\mathbb{P}_X(x) = \widehat{\mathbb{P}}_X(t)$, on constate que la fonction caractéristique n'est rien d'autre que la transformée de Fourier de la loi de X . Cette fonction est très utile car elle détermine entièrement la loi de X . Nous vous renvoyons vers le cours d'analyse harmonique pour plus de détails.

Théorème 2.6.4 L'application $\mathbb{P}_X \mapsto \varphi^X$ définie dans l'espace des mesures de probabilité sur \mathbb{R}^d est injective.

Corollaire 2.6.5 Deux vecteurs aléatoires sur \mathbb{R}^d sont de même loi si et seulement si ils ont la même fonction caractéristique.

Remarque 2.6.6 Les faits suivants sont utiles.

- Si $X = a$ p.s., alors $\varphi^X(t) = \exp(it^T a)$.
- Si X est un vecteur aléatoire dans \mathbb{R}^d , si A est une matrice agissant sur \mathbb{R}^d et si $m \in \mathbb{R}^d$, alors

$$\varphi^{AX+m} = \exp(it^T m) \varphi^X(A^T t).$$

Proposition 2.6.7 Soit $X = (X_1, \dots, X_d)^T$ un vecteur aléatoire dans \mathbb{R}^d .

- Si $X \in L^1$, alors φ^X est de classe C^1 et $\nabla \varphi^X(0) = i\mathbb{E}[X]^T$.
- Si $X \in L^2$, alors φ^X est de classe C^2 et $\text{Hess}_{\varphi^X}(0) = -\mathbb{E}[XX^T]$. De plus

$$\varphi^X(t) = 1 + i \sum_{j=1}^d t_j \mathbb{E}[X_j] - \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d t_j t_k \mathbb{E}[X_j X_k] + o(|t|^2)$$

quand $t = (t_1, \dots, t_d)$ tend vers 0.

Démonstration : Pour le premier point, en dérivant sous le signe intégrale, on trouve

$$\frac{\partial \varphi^X}{\partial t_j}(t) = i \mathbb{E} \left[X_j e^{it^T X} \right],$$

l'opération étant permise puisque $|iX_j e^{it^T X}| = |X_j|$ et $X_j \in L^1$. Pour le deuxième point, puisque $\mathbb{E}[|X_j X_k|] \leq \mathbb{E}[X_j^2]^{1/2} \mathbb{E}[X_k^2]^{1/2} < +\infty$ par Cauchy-Schwarz, on peut dériver une seconde fois et trouver que

$$\frac{\partial^2 \varphi^X}{\partial t_j \partial t_k}(t) = -\mathbb{E} \left[X_j X_k e^{it^T X} \right].$$

De plus le théorème de continuité sous le signe intégrale nous assure que cette fonction est continue. Enfin la dernière affirmation est simplement le développement de Taylor à l'ordre 2. ■

En dimension $d = 1$ on peut pousser l'argument précédent à n'importe quel ordre en utilisant à nouveau le théorème de dérivation des intégrales paramétriques.

Proposition 2.6.8 Soit X une variable aléatoire de fonction caractéristique φ . Si $X \in L^n$, alors φ est de classe C^n et

$$\varphi^{(k)}(t) = i^k \mathbb{E} \left[X^k e^{itX} \right]$$

pour tout $k \leq n$. En particulier, $\varphi^{(k)}(0) = i^k \mathbb{E}[X^k]$.

Il est commun d'introduire une fonction qui ressemble fort à la fonction caractéristique, à savoir la fonction génératrice des moments.

Définition 2.6.9 Soit X un vecteur aléatoire dans \mathbb{R}^d . Sa *fonction génératrice des moments* (aka transformée de Laplace) est la fonction

$$L^X(t) = \mathbb{E} \left[e^{t^T X} \right]$$

définie en les valeurs de t en lesquelles $e^{t^T X} \in L^1$.

Il s'agit ici de la transformée de Laplace de la loi de X . Contrairement à la fonction caractéristique, la fonction génératrice des moments n'est pas nécessairement définie; afin de pouvoir l'utiliser il faut supposer que la fonction $e^{t^T X}$ est intégrable dans un voisinage de 0. Bien entendu, $L_X(0) = 1$. On peut dire beaucoup plus. Supposons que X est une variable aléatoire réelle qui admet une fonction génératrice des moments pour $t > 0$, et supposons de plus que nous pouvons échanger dérivées et espérance. Alors

$$\begin{aligned} \mathbb{E}[e^{tX}] &= \mathbb{E} \left[1 + tX + \frac{1}{2}t^2X^2 + \frac{1}{3!}t^3X^3 + \dots \right] \\ &= 1 + t\mathbb{E}[X] + \frac{1}{2}t^2\mathbb{E}[X^2] + \frac{1}{3!}t^3\mathbb{E}[X^3] + \dots \end{aligned}$$

De cette réécriture (formelle), on a le résultat suivant qui justifie la terminologie.

Proposition 2.6.10 Soit X une variable aléatoire telle que e^{tX} est intégrable pour t dans un intervalle ouvert contenant 0. Alors, la fonction génératrice des moments L^X de X est définie et analytique sur un intervalle ouvert contenant 0. On a

$$L^X(t) = \sum_{n \in \mathbb{N}} \frac{t^n}{n!} \mathbb{E}[X^n]$$

pour tout t dans ce voisinage. En particulier, pour tout $n \in \mathbb{N}$,

$$(L^X)^{(n)}(0) = \mathbb{E}[X^n].$$

Remarque 2.6.11 Si la fonction génératrice des moments de X existe, elle caractérise une infinité de moments de X . Ceci ne suffit toutefois pas à caractériser la loi de X : on peut construire X_1 et X_2 qui ont deux lois différentes mais telles que $\mathbb{E}[X_1^n] = \mathbb{E}[X_2^n]$ pour tout n . C'est le cas des variables aléatoires continues X_1 et X_2 de densités

$$f_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2} \mathbf{1}_{[0,+\infty[}(x)$$

$$f_2(x) = f_1(x) [1 + \sin(2\pi \log x)] \mathbf{1}_{[0,+\infty[}(x)$$

respectivement. On peut montrer que $\mathbb{E}[X_1^n] = \mathbb{E}[X_2^n] = e^{n^2/2}$ mais toutefois X_1 et X_2 n'ont pas la même loi. Néanmoins, dans le cas où X est une variable aléatoire *bornée*, sa loi est entièrement déterminée par ses moments. Ce résultat s'appelle le *théorème des moments*.

On admettra toutefois le résultat suivant.

Proposition 2.6.12 Soient X, Y deux variables aléatoires telles qu'il existe un voisinage I de 0 sur lequel L_X et L_Y sont définies. Alors, $X \stackrel{\mathcal{L}}{=} Y$ si et seulement si $L_X = L_Y$ sur I .

Chapitre 3

Indépendance

3.1 Indépendance d'événements

C'est avec la notion d'indépendance que la théorie des probabilités prend son autonomie par rapport à la théorie de la mesure. Nous rappelons la définition de l'indépendance d'événements.

Définition 3.1.1 Soient $A_1, \dots, A_n \in \mathcal{F}$. On dit que des événements A_1, \dots, A_n sont *indépendants* si pour toute partie finie non-vide J de $\{1, \dots, n\}$, on a

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Le résultat suivant montre que si des événements A_1, \dots, A_n sont indépendants, alors on peut remplacer des événements par leur complémentaire et garder une famille d'événements indépendants. De manière générale, on l'écrit en termes de la σ -algèbre engendrée par un événement. Cette propriété permettra d'étendre la définition à l'indépendance de σ -algèbres.

Lemme 3.1.2 Si $A_1, \dots, A_n \in \mathcal{F}$ sont indépendants, alors A_1^c, A_2, \dots, A_n sont également indépendants.

Démonstration : On peut bien sûr considérer une partie J telle que $1 \in J$. On remarque alors que

$$\bigcap_{j \in J, j \neq 1} A_j = \left(A_1^c \cap \bigcap_{j \in J, j \neq 1} A_j \right) \cup \bigcap_{j \in J} A_j$$

et puisque cette union est disjointe, il vient

$$\begin{aligned} \mathbb{P}\left(A_1^c \cap \bigcap_{j \in J, j \neq 1} A_j\right) &= \mathbb{P}\left(\bigcap_{j \in J, j \neq 1} A_j\right) - \mathbb{P}\left(\bigcap_{j \in J} A_j\right) \\ &= \prod_{j \in J, j \neq 1} \mathbb{P}(A_j) - \prod_{j \in J} \mathbb{P}(A_j) \\ &= (1 - \mathbb{P}(A_1)) \prod_{j \in J, j \neq 1} \mathbb{P}(A_j) \\ &= \mathbb{P}(A_1^c) \prod_{j \in J, j \neq 1} \mathbb{P}(A_j). \end{aligned}$$

Proposition 3.1.3 Soient $A_1, \dots, A_n \in \mathcal{F}$. Les événements A_1, \dots, A_n sont indépendants si et seulement si

$$\mathbb{P}\left(\bigcap_{j=1}^n B_j\right) = \prod_{j=1}^n \mathbb{P}(B_j).$$

où pour tout $j \in \{1, \dots, n\}$ on a $B_j \in \sigma(A_j) = \{\emptyset, A_j, A_j^c, \Omega\}$.

Démonstration : Il est clair que la condition donnée est plus forte que la définition : il suffit de prendre $B_j = A_j$ si $j \in J$ et $B_j = \Omega$ sinon. Pour vérifier la réciproque, on peut bien sûr supposer que $B_j \neq \emptyset$ pour tout j . Si à présent $\{j : B_j \neq \Omega\} = \{j_1, \dots, j_p\}$, il suffit de montrer que

$$\mathbb{P}(B_{j_1} \cap \dots \cap B_{j_p}) = \mathbb{P}(B_{j_1}) \dots \mathbb{P}(B_{j_p}).$$

Pour cela, il suffit de procéder de proche en proche en appliquant au plus p fois le lemme précédent. ■

Terminons cette courte section par le lemme de Borel-Cantelli qui permet de calculer la probabilité d'événements limites. Il s'agit d'un résultat général de théorie de la mesure (cfr Proposition A.3.9). Néanmoins, la notion d'indépendance permet d'obtenir une réciproque à ce résultat. Rappelons que si $(A_n)_{n \in \mathbb{N}}$ est une suite d'ensembles, alors

$$\limsup_{n \rightarrow +\infty} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \quad \text{et} \quad \liminf_{n \rightarrow +\infty} A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m.$$

Proposition 3.1.4 (Lemme de Borel-Cantelli) Soit $(A_n)_{n \in \mathbb{N}}$ une suite de \mathcal{F} .

1. Si

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < +\infty,$$

alors

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0.$$

2. Si

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$$

et si les événements A_n , $n \in \mathbb{N}$, sont indépendants, alors

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 1.$$

Démonstration : 1. Pour tout $n \in \mathbb{N}$, on pose $B_n = \bigcup_{k \geq n} A_k$. Alors $(B_n)_{n \in \mathbb{N}}$ est une suite décroissante d'éléments de \mathcal{F} . La continuité à droite de la mesure \mathbb{P} donne

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) \leq \lim_{n \rightarrow +\infty} \sum_{k \geq n} \mathbb{P}(A_k) = 0.$$

2. Comme dans le point précédent, la continuité à droite de la mesure \mathbb{P} implique qu'il suffit de montrer que

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\bigcup_{k \geq n} A_k \right) = 1,$$

ou de manière équivalente, que

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) = 0.$$

Fixons $n \in \mathbb{N}$ et remarquons que par indépendance des événements, on a

$$\mathbb{P} \left(\bigcap_{k=n}^m A_k^c \right) = \prod_{k=n}^m \mathbb{P}(A_k^c) = \prod_{k=n}^m (1 - \mathbb{P}(A_k))$$

si $m \geq n$. La continuité à droite de \mathbb{P} implique que

$$\mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) = \lim_{m \rightarrow +\infty} \prod_{k=n}^m (1 - \mathbb{P}(A_k)).$$

Or, pour tout $x \in [0, 1]$, on a $(1 - x) \leq e^{-x}$ et donc

$$\prod_{k=n}^m (1 - \mathbb{P}(A_k)) \leq \prod_{k=n}^m e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^m \mathbb{P}(A_k)} = e^{-\sum_{k=n}^m \mathbb{P}(A_k)}$$

qui, par hypothèse, tend vers 0 lorsque m tend vers l'infini. ■

La lemme de Borel-Cantelli signifie donc que

1. si la série $\sum_n \mathbb{P}(A_n)$ converge, alors il est presque sûr qu'à partir d'un certain rang, l'événement A_n n'a pas lieu,
2. si la série $\sum_n \mathbb{P}(A_n)$ diverge et si les événements sont indépendants, alors il est presque sûr que l'événement A_n a lieu pour une infinité de n .

Exemple 3.1.5 On considère l'expérience qui consiste à lancer une infinité de fois une pièce de monnaie. On note A_n l'événement "Le $n^{\text{ème}}$ lancé est Pile". Comme $\mathbb{P}(A_n) = 1/2$ et que les variables aléatoires A_n , $n \geq 1$ sont indépendantes, la réciproque du lemme de Borel-Cantelli implique que $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$. Autrement dit, presque sûrement, on aura une infinité de fois Pile.

Remarque 3.1.6 On ne peut pas se passer de l'hypothèse d'indépendance. Par exemple, si on se place sur l'espace $\Omega = [0, 1]$ muni de la mesure de Lebesgue $\mathbb{P} = \lambda|_{[0,1]}$, et si on considère les événements $A_n = [0, 1/n]$, $n \in \mathbb{N}_0$, alors on a $\sum_n \mathbb{P}(A_n) = \sum_n \frac{1}{n} = +\infty$ mais $\limsup_n A_n = \{0\}$.

3.2 Indépendance de σ -algèbres

Dans cette section, nous généralisons la notion d'indépendance d'événements aux cas de collections d'événements, et en particulier de sous- σ -algèbres.

Définition 3.2.1

- Une famille $A_i, i \in I$, d'événements de \mathcal{F} est *indépendante* si toute sous-famille finie est formée d'événements indépendants. On dit également que les événements $A_i, i \in I$, sont indépendants.
- Une famille quelconque de collections d'événements $\mathcal{C}_i \subseteq \mathcal{F}, i \in I$, est *indépendante* si toute famille d'événements $A_i \in \mathcal{C}_i, i \in I$, est indépendante. On dit également que les collections $\mathcal{C}_i, i \in I$, sont indépendantes.

Dans le cas où les collections \mathcal{C}_i sont des sous- σ -algèbres de \mathcal{F} , on a le critère pratique suivant.

Proposition 3.2.2 Soient $\mathcal{F}_1, \dots, \mathcal{F}_n$ des sous- σ -algèbres de \mathcal{F} . Les collections $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes si et seulement si

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n)$$

pour tous $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$.

Démonstration : Il est clair que la relation est vérifiée si les sous- σ -algèbres sont indépendantes. Réciproquement, si on a cette propriété et si J est une partie de $\{1, \dots, n\}$, il suffit de prendre $A_j = \Omega$ si $j \notin J$ pour obtenir l'égalité de la Définition B.3.8. ■

Remarquons donc que la Proposition 3.1.3 signifie exactement que des événements A_1, \dots, A_n sont indépendants si et seulement si les σ -algèbres $\sigma(A_1), \dots, \sigma(A_n)$ sont indépendantes.

Le résultat suivant permet de montrer que pour vérifier l'indépendance de σ -algèbres, il suffit de le vérifier sur des classes stables par intersection finie qui les engendrent : ce résultat est semblable au lemme de la classe monotone et se démontre facilement à partir de celui-ci.

Proposition 3.2.3 Soit $\mathcal{C}_i, i \in I$, des collections d'événements de \mathcal{F} . Si chaque \mathcal{C}_i est stable par intersection finie, alors les collections $\mathcal{C}_i, i \in I$, sont indépendantes si et seulement si les σ -algèbres $\sigma(\mathcal{C}_i), i \in I$, sont indépendantes.

Démonstration : Evidemment, si les σ -algèbres $\sigma(\mathcal{C}_i), i \in I$, sont indépendantes, alors les collections $\mathcal{C}_i, i \in I$, le sont également. Supposons donc que les $\mathcal{C}_i, i \in I$, sont indépendants. Soit $J \subseteq I$ de cardinalité finie n . Sans perte de généralité, on suppose que $J = \{1, \dots, n\}$. Pour tout $j \in \{2, \dots, n\}$, soit $A_j \in \mathcal{C}_j$. Notons μ_1 et μ_2 les mesures finies définies par

$$\mu_1 : \sigma(\mathcal{C}_1) \rightarrow [0, 1] : A_1 \mapsto \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n)$$

et

$$\mu_2 : \sigma(\mathcal{C}_1) \rightarrow [0, 1] : A_1 \mapsto \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n).$$

Par hypothèse, les mesures μ_1 et μ_2 sont égales sur \mathcal{C}_1 et le lemme de la classe monotone implique donc que $\mu_1 = \mu_2$ sur $\sigma(\mathcal{C}_1)$. On a donc $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n)$ pour tout $A_1 \in \sigma(\mathcal{C}_1)$ et tous $A_2 \in \mathcal{C}_2, \dots, A_n \in \mathcal{C}_n$. On recommence ensuite en fixant $A_1 \in \sigma(\mathcal{C}_1)$ et $A_3 \in \mathcal{C}_3, \dots, A_n \in \mathcal{C}_n$ pour montrer que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n)$ pour tout $A_1 \in \sigma(\mathcal{C}_1), A_2 \in \sigma(\mathcal{C}_2)$ et tous $A_3 \in \mathcal{C}_3, \dots, A_n \in \mathcal{C}_n$. On continue ainsi de suite pour obtenir la conclusion. ■

Le résultat suivant, très intuitif, montre qu'en réunissant des σ -algèbres indépendantes, on construit des collections indépendantes.

Théorème 3.2.4 (Indépendance par paquets) Soit $\mathcal{F}_i, i \in I$, une collection indépendante de sous- σ -algèbres de \mathcal{F} . Si $I_j, j \in J$, sont des sous-ensembles deux à deux disjoints de I , alors les σ -algèbres $\sigma(\bigcup_{i \in I_j} \mathcal{F}_i), j \in J$, sont indépendantes.

Démonstration : Pour tout $j \in J$, on note

$$\mathcal{C}_j = \left\{ A_1 \cap \dots \cap A_n : n \in \mathbb{N}, A_1, \dots, A_n \in \bigcup_{i \in I_j} \mathcal{F}_i \right\}.$$

Clairement, \mathcal{C}_j est stable par intersection finie. De plus, il est facile de voir que $\mathcal{C}_j \subseteq \sigma\left(\bigcup_{i \in I_j} \mathcal{F}_i\right)$. D'autre part, comme $\mathcal{F}_i \subseteq \mathcal{C}_j$ pour tout $i \in I_j$, on a également l'autre inclusion, d'où

$$\sigma(\mathcal{C}_j) = \sigma\left(\bigcup_{i \in I_j} \mathcal{F}_i\right).$$

La conclusion résultera du Théorème 3.2.3 si on montre que les collections $\mathcal{C}_j, j \in J$, sont indépendantes. Soient donc j_1, \dots, j_r des indices différents de J et considérons $B_k \in \mathcal{C}_{j_k}$ pour tout $k \in \{1, \dots, r\}$. Alors, pour tout $k \in \{1, \dots, r\}$, on peut écrire B_k sous la forme

$$B_k = \bigcap_{m \in K_k} A_m$$

où $K_k \subseteq I_{j_k}$ est fini et $A_m \in \mathcal{F}_m$. Alors, les ensembles $K_k, k \in \{1, \dots, r\}$ sont disjoints et on a

$$\mathbb{P}\left(\bigcap_{k=1}^r B_k\right) = \mathbb{P}\left(\bigcap_{k=1}^r \bigcap_{m \in K_k} A_m\right) = \prod_{k=1}^r \prod_{m \in K_k} \mathbb{P}(A_m) = \prod_{k=1}^r \mathbb{P}(B_k),$$

ce qui suffit. ■

Exemple 3.2.5 Si $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements indépendants, alors les événements $A_1, A_2 \cup A_3, A_4 \cap A_6, A_5^c, \bigcup_{n \geq 10} A_n$ sont également indépendants.

Nous terminons cette section par un résultat très surprenant mais également très important concernant les événements "limites".

Définition 3.2.6 Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une suite de sous- σ -algèbres de \mathcal{F} . La σ -algèbre

$$\mathcal{F}_\infty = \bigcap_{n \in \mathbb{N}} \sigma\left(\bigcup_{j \geq n} \mathcal{F}_j\right)$$

est appelée la σ -algèbre asymptotique de la suite $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

Etant une intersection de σ -algèbres, il est clair que \mathcal{F}_∞ est également une σ -algèbre. Intuitivement, les événements de \mathcal{F}_∞ décrivent ce qui se passe à l'infini, relativement à la suite $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

Exemple 3.2.7 Si $A_n \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$, alors $\limsup_n A_n \in \mathcal{F}_\infty$ et $\liminf_n A_n \in \mathcal{F}_\infty$. Fixons $n \in \mathbb{N}$ et montrons que $\liminf_n A_n \in \sigma\left(\bigcup_{j \geq n} \mathcal{F}_j\right)$. En effet, pour tout $m \geq n$, on a $\bigcap_{k \geq m} A_k \in \sigma\left(\bigcup_{j \geq n} \mathcal{F}_j\right)$ et comme la suite $\left(\bigcap_{k \geq m} A_k\right)_{m \in \mathbb{N}}$ est croissante, on a

$$\liminf_{n \rightarrow +\infty} A_n = \bigcup_{m \in \mathbb{N}} \bigcap_{k \geq m} A_k = \bigcup_{m \geq n} \bigcap_{k \geq m} A_k \in \sigma\left(\bigcup_{j \geq n} \mathcal{F}_j\right).$$

Le résultat pour la limite supérieure s'en déduit par passage au complémentaire.

Nous savons que si les événements A_n , $n \in \mathbb{N}$, sont indépendants, alors l'ensemble $\limsup_n A_n$ est de probabilité 1 ou 0 selon que la série de terme principal $\mathbb{P}(A_n)$ converge ou diverge : c'est le lemme de Borel-Cantelli. La loi du 0-1 de Kolmogorov montre que le résultat est vrai pour tout événement de la σ -algèbre asymptotique.

Théorème 3.2.8 (Loi du 0-1 de Kolmogorov) *Si les sous- σ -algèbres $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de \mathcal{F} sont indépendantes, alors pour tout $A \in \mathcal{F}_\infty$, on a*

$$\mathbb{P}(A) = 0 \quad \text{ou} \quad \mathbb{P}(A) = 1.$$

Démonstration : Nous allons montrer que la σ -algèbre \mathcal{F}_∞ est indépendant d'elle-même. Ainsi, si $A \in \mathcal{F}_\infty$, alors $A \perp\!\!\!\perp A$ et donc $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ d'où la conclusion.

Comme les sous- σ -algèbres $(\mathcal{F}_n)_{n \in \mathbb{N}}$ sont indépendantes, le Théorème 3.2.4 implique que pour tout $n \in \mathbb{N}_0$, les σ -algèbres $\sigma(\bigcup_{j < n} \mathcal{F}_j)$ et $\sigma(\bigcup_{j \geq n} \mathcal{F}_j)$ sont indépendantes. Puisque $\mathcal{F}_\infty \subseteq \sigma(\bigcup_{j \geq n} \mathcal{F}_j)$, on en tire que pour tout $n \in \mathbb{N}_0$, les σ -algèbres $\sigma(\bigcup_{j < n} \mathcal{F}_j)$ et \mathcal{F}_∞ sont indépendantes. Ainsi, pour tout $n \in \mathbb{N}_0$, tout $A \in \sigma(\bigcup_{j < n} \mathcal{F}_j)$ et tout $B \in \mathcal{F}_\infty$, on a

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

On en tire que les collections $\mathcal{C} = \bigcup_{n \in \mathbb{N}_0} \sigma(\bigcup_{j < n} \mathcal{F}_j)$ et \mathcal{F}_∞ sont indépendantes. Comme \mathcal{C} est une union de σ -algèbres croissantes, \mathcal{C} est stable par intersection finie. Dès lors, la Proposition 3.2.3 implique que les σ -algèbres $\sigma(\mathcal{C})$ et \mathcal{F}_∞ sont indépendantes. Pour conclure, il suffit maintenant de remarquer que $\mathcal{F}_\infty \subseteq \sigma(\mathcal{C})$: en effet, $\mathcal{F}_j \subseteq \mathcal{C}$ pour tout $j \in \mathbb{N}$, d'où $\bigcup_{j \in \mathbb{N}} \mathcal{F}_j \subseteq \mathcal{C}$ et donc $\sigma(\bigcup_{j \in \mathbb{N}} \mathcal{F}_j) \subseteq \sigma(\mathcal{C})$, ce qui suffit puisque $\mathcal{F}_\infty \subseteq \sigma(\bigcup_{j \in \mathbb{N}} \mathcal{F}_j)$. ■

3.3 Indépendance de variables aléatoires

L'indépendance des variables aléatoires se définit à partir des σ -algèbres associées. Rappelons que si $X : \Omega \rightarrow \mathbb{R}$, il existe une plus petite σ -algèbre sur Ω qui rend X mesurable. C'est la σ -algèbre engendrée par X , que l'on note $\sigma(X)$. La Proposition A.2.8 permet d'écrire

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}.$$

C'est donc la σ -algèbre qui porte toute l'information nécessaire à la connaissance de X . Rappelons que si X est une variable aléatoire, alors $\sigma(X) \subseteq \mathcal{F}$.

Définition 3.3.1 Soit $X_i, i \in I$, une famille de variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que les variables aléatoires $X_i, i \in I$, sont *indépendantes* si les sous- σ -algèbres $\sigma(X_i), i \in I$, sont indépendantes.

La description de $\sigma(X)$ mène directement à la caractérisation pratique suivante.

Proposition 3.3.2 Des variables aléatoires $X_i, i \in I$, sont indépendantes si et seulement si pour tout $n \in \mathbb{N}_0$, tous $B_1, \dots, B_n \in \mathcal{B}$ et tous $j_1, \dots, j_n \in I$ distincts, on a

$$\mathbb{P}(X_{j_1} \in B_1, \dots, X_{j_n} \in B_n) = \mathbb{P}(X_{j_1} \in B_1) \dots \mathbb{P}(X_{j_n} \in B_n).$$

Exemple 3.3.3 On lance n fois une pièce. Pour tout $j \in \{1, \dots, n\}$, on considère la variable aléatoire X_j qui vaut 1 si le $j^{\text{ème}}$ lancer est “pile” et 0 sinon. Alors, les variables aléatoires X_1, \dots, X_n sont indépendantes. Intuitivement, c’est clair. Mathématiquement, on le démontre comme suit. Soit $\Omega = \{0, 1\}^n$ muni de la mesure de probabilité uniforme \mathbb{P} (chaque issue est équiprobable). Les variables aléatoires X_j sont définies par $X_j(\omega_1, \dots, \omega_n) = \omega_j$. Soient $A_1, \dots, A_n \subseteq \{0, 1\}$. On a

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \frac{\#(A_1 \times \dots \times A_n)}{\#\Omega} = \frac{\#A_1 \dots \#A_n}{2^n} = \frac{\#A_1}{2} \dots \frac{\#A_n}{2}$$

et d’autre part, $\{X_j \in A_j\} = \{0, 1\} \times \dots \times A_j \times \dots \times \{0, 1\}$ d’où

$$\mathbb{P}(X_j \in A_j) = \frac{\#\{0, 1\} \times \dots \times A_j \times \dots \times \{0, 1\}}{\#\Omega} = \frac{2^{n-1} \#A_j}{2^n} = \frac{\#A_j}{2}.$$

La conclusion s’ensuit.

Exemple 3.3.4 Soit $\Omega = [0, 1] \times [0, 1]$ muni de la mesure de Lebesgue $\mathbb{P} = \lambda_2$ sur Ω . Soient X et Y les variables aléatoires définies sur Ω par $X(x, y) = x$ et $Y(x, y) = y$. Alors, pour tous boréliens $A, B \in \mathcal{B}$, on a $\{X \in A, Y \in B\} = A \times B$ d’où

$$\mathbb{P}(X \in A, Y \in B) = \lambda_2(A \times B) = \lambda_1(A) \lambda_1(B) = \lambda_2(A \times [0, 1]) \lambda_2([0, 1] \times B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B).$$

Les variables aléatoires X et Y sont donc indépendantes.

Proposition 3.3.5 Des variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si la loi du vecteur aléatoire (X_1, \dots, X_n) se factorise en le produit des lois des marginales, i.e.

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_n}.$$

Démonstration : Remarquons que si $B_1, \dots, B_n \in \mathcal{B}$, alors d’une part on a

$$\mathbb{P}_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) = \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n)$$

et d’autre part

$$\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_n}(B_1 \times \dots \times B_n) = \mathbb{P}_{X_1}(B_1) \dots \mathbb{P}_{X_n}(B_n) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)$$

par définition de la mesure produit. La Proposition 3.3.2 implique que les mesures $\mathbb{P}_{(X_1, \dots, X_n)}$ et $\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_n}$ sont égales sur les rectangles mesurables si et seulement si les variables aléatoires X_1, \dots, X_n sont indépendantes. Le lemme de la classe monotone permet d’affirmer que les mesures sont alors égales sur \mathcal{B}^n . ■

Remarque 3.3.6 Le résultat précédent montre également comment construire des variables aléatoires indépendantes : soient μ_1, \dots, μ_n des mesures de probabilités sur \mathbb{R} . Alors, on peut considérer la mesure produit $\mu_1 \times \dots \times \mu_n$ sur \mathbb{R}^n et définir un vecteur aléatoire dont la loi est donnée par $\mu = \mu_1 \times \dots \times \mu_n$. Il suffit de prendre $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}^n$, $\mathbb{P} = \mu_1 \times \dots \times \mu_n$ et de poser $X(\omega) = \omega$. La loi de X est μ et les composantes de X sont des variables aléatoires indépendantes de loi μ_1, \dots, μ_n respectivement.

Regardons à présent ce que la Proposition 3.3.2 implique sur les fonctions de répartition, sur les fonctions de masse dans le cas discret et fonctions de densité dans le cas continu.

Proposition 3.3.7 *Des variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si*

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$$

pour tous $x_1, \dots, x_n \in \mathbb{R}$.

Démonstration : D'une part, on a

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}_{(X_1, \dots, X_n)}([\!-\infty, x_1] \times \dots \times]-\infty, x_n])$$

et d'autre part

$$F_{X_1}(x_1) \dots F_{X_n}(x_n) = \mathbb{P}(X_1 \leq x_1) \dots \mathbb{P}(X_n \leq x_n) = \mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_n}([\!-\infty, x_1] \times \dots \times]-\infty, x_n]).$$

Si les variables aléatoires sont indépendantes, la conclusion provient alors de la Proposition 3.3.2. Réciproquement, le lemme de la classe monotone donne la conclusion puisque par hypothèse, les mesures $\mathbb{P}_{(X_1, \dots, X_n)}$ et $\mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_n}$ coïncident sur les demi-espaces. ■

Dans le cas où les variables aléatoires sont discrètes, il suffit de vérifier que la fonction de masse de la distribution jointe est égale au produit des fonctions de masse marginales.

Proposition 3.3.8 *Soient X_1, \dots, X_n des variables aléatoires discrètes. Les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si pour tous $x_1, \dots, x_n \in \mathbb{R}$, on a*

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n).$$

Démonstration : Si les variables aléatoires sont indépendantes, la conclusion découle directement de la Proposition 3.3.2. Réciproquement, si les valeurs prises par X_1, \dots, X_n sont données par les ensembles $\Lambda_1, \dots, \Lambda_n$ respectivement, l'hypothèse implique que

$$\begin{aligned} \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) &= \sum_{y_1 \in \Lambda_1: y_1 \leq x_1} \dots \sum_{y_n \in \Lambda_n: y_n \leq x_n} \mathbb{P}(X_1 = y_1, \dots, X_n = y_n) \\ &= \sum_{y_1 \in \Lambda_1: y_1 \leq x_1} \dots \sum_{y_n \in \Lambda_n: y_n \leq x_n} \mathbb{P}(X_1 = y_1) \dots \mathbb{P}(X_n = y_n) \\ &= F_{X_1}(x_1) \dots F_{X_n}(x_n), \end{aligned}$$

ce qui suffit par la Proposition 3.3.7. ■

Dans le cas continu, on a naturellement que l'indépendance des variables aléatoires est équivalente à la factorisation de la densité de la distribution jointe en le produit des densités marginales.

Proposition 3.3.9 Soient X_1, \dots, X_n des variables aléatoires continues. Si le vecteur aléatoire (X_1, \dots, X_n) est continu, alors les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n)$$

pour presque tout $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Démonstration : Si $f_{(X_1, \dots, X_n)} = f_{X_1} \dots f_{X_n}$, on a

$$\begin{aligned} F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{(X_1, \dots, X_n)}(y_1, \dots, y_n) dy_1 \dots dy_n \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1}(y_1) \dots f_{X_n}(y_n) dy_1 \dots dy_n \\ &= \int_{-\infty}^{x_1} f_{X_1}(y_1) dy_1 \dots \int_{-\infty}^{x_n} f_{X_n}(y_n) dy_n \\ &= F_{X_1}(x_1) \dots F_{X_n}(x_n) \end{aligned}$$

ce qui suffit par la Proposition 3.3.7. Réciproquement, si les variables aléatoires sont indépendantes, alors $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \times \dots \times \mathbb{P}_{X_n}$ par la Proposition 3.3.5. Par unicité de la densité associée à une mesure, on en tire que $f_{(X_1, \dots, X_n)} = f_{X_1} \dots f_{X_n}$ presque partout. ■

Remarque 3.3.10 Le résultat précédent montre même que si les variables aléatoires X_1, \dots, X_n sont continues et indépendantes, alors le vecteur aléatoire (X_1, \dots, X_n) est également continu (ce qui est en général faux).

Bien souvent, on ne démontre pas que des variables aléatoires sont indépendantes mais on prend ce fait comme hypothèse. Le résultat ci-dessous en est un exemple et montre que, sous l'hypothèse d'indépendance des variables aléatoires, le minimum de deux variables aléatoires suivant une loi géométrique suit encore une loi géométrique.

Proposition 3.3.11 (Stabilité par les extrêmes) Si X et Y sont des variables aléatoires indépendantes de loi géométrique de paramètres p et p' respectivement, alors $\min(X, Y)$ est une variable aléatoire de loi géométrique de paramètre $1 - (1 - p)(1 - p')$.

Démonstration : Il est clair que $\min(X, Y)$ est à valeurs dans \mathbb{N}_0 . De plus, pour tout $k \in \mathbb{N}_0$, on a

$$\begin{aligned} \mathbb{P}(\min(X, Y) > k) &= \mathbb{P}(X > k, Y > k) \\ &= \mathbb{P}(X > k) \mathbb{P}(Y > k) \\ &= (1 - p)^k (1 - p')^k = ((1 - p)(1 - p'))^k \end{aligned}$$

en utilisant la Proposition 1.4.6 et l'indépendance. La conclusion suit. ■

Dans les sections suivantes, nous verrons des propriétés de stabilité de loi par rapport à la somme de variables aléatoires indépendantes.

3.4 Loi du 0-1 de Kolmogorov

Dans cette section, nous donnons quelques conséquences de la loi du 0-1 de Kolmogorov dans le langage des variables aléatoires.

Si on considère une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires, on peut naturellement lui associer une suite $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous- σ -algèbres de \mathcal{F} telle que X_n est \mathcal{F}_n -mesurable pour tout $n \in \mathbb{N}$: il suffit de poser $\mathcal{F}_n = \sigma(X_n)$. De plus, rappelons que les variables aléatoires X_n , $n \in \mathbb{N}$ sont indépendantes si les σ -algèbres $\sigma(X_n)$, $n \in \mathbb{N}$ le sont. La σ -algèbre asymptotique des $\sigma(X_n)$, $n \in \mathbb{N}$ contient tous les événements qu'on peut définir à partir de la suite $(X_n)_{n \in \mathbb{N}}$ en ignorant n'importe quelle sous-collection finie de variables. Ce sont des événements que l'on rencontre très régulièrement.

Proposition 3.4.1 *Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une suite de sous- σ -algèbres de \mathcal{F} et soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Si X_n est \mathcal{F}_n -mesurable pour tout $n \in \mathbb{N}$, alors $\limsup_{n \rightarrow +\infty} X_n$ et $\liminf_{n \rightarrow +\infty} X_n$ est \mathcal{F}_∞ -mesurable.*

Démonstration : Il suffit de montrer le résultat pour $\liminf_{n \rightarrow +\infty} X_n$. On remarque que la variable aléatoire $Y_n := \inf_{k \geq n} X_k$ est $\sigma(\bigcup_{k \geq n} \mathcal{F}_k)$ -mesurable. Par conséquent, $\sup_{n \geq m} Y_n$ est $\sigma(\bigcup_{n \geq m} \mathcal{F}_n)$ -mesurable. Or, pour tout $m \in \mathbb{N}$, on a $\liminf_{n \rightarrow +\infty} X_n = \sup_{n \geq m} Y_n$ et donc $\liminf_{n \rightarrow +\infty} X_n$ est $\sigma(\bigcup_{n \geq m} \mathcal{F}_n)$ -mesurable pour tout $m \in \mathbb{N}$, et donc \mathcal{F}_∞ -mesurable. ■

Remarque 3.4.2 On considère ici les variables aléatoires $\limsup_{n \rightarrow +\infty} X_n$ et $\liminf_{n \rightarrow +\infty} X_n$ à valeurs dans $\overline{\mathbb{R}}$.

Corollaire 3.4.3 *Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une suite de sous- σ -algèbres de \mathcal{F} et soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Si X_n est \mathcal{F}_n -mesurable pour tout $n \in \mathbb{N}$, alors l'ensemble*

$$\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) \text{ existe}\}$$

appartient à \mathcal{F}_∞ .

Démonstration : C'est immédiat par le résultat précédent puisque

$$\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) \text{ existe}\} = \{\omega \in \Omega : \liminf_{n \rightarrow +\infty} X_n(\omega) = \limsup_{n \rightarrow +\infty} X_n(\omega)\}.$$

On sait que les variables aléatoires X_n , $n \in \mathbb{N}$ sont indépendantes si les σ -algèbres $\sigma(X_n)$, $n \in \mathbb{N}$ le sont. La loi du 0-1 (aussi appelée la loi du tout ou du rien) dans le contexte des variables aléatoires (c'est-à-dire pour les σ -algèbres engendrées par les X_n) donne le résultat suivant.

Corollaire 3.4.4 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes. Alors, les variables aléatoires $\limsup_{n \rightarrow +\infty} X_n$ et $\liminf_{n \rightarrow +\infty} X_n$ sont presque sûrement constantes.*

Démonstration : On considère les σ -algèbres indépendantes $\mathcal{F}_n = \sigma(X_n)$, $n \in \mathbb{N}$. On sait déjà que $X := \liminf_{n \rightarrow +\infty} X_n$ est \mathcal{F}_∞ -mesurable. Si $\mathbb{P}(X = -\infty) = 1$ ou $\mathbb{P}(X = +\infty) = 1$, alors $X = -\infty$ ou $X = +\infty$ respectivement presque sûrement. Sinon, la loi du 0-1 implique que $\mathbb{P}(X = -\infty) = \mathbb{P}(X = +\infty) = 0$ et donc X est presque sûrement à valeurs réelles. De plus, pour tout $x \in \mathbb{R}$, on a

$$F_X(x) = \mathbb{P}(X \leq x) \in \{0, 1\}$$

et donc si on pose $a = \sup\{x \in \mathbb{R} : F_X(x) = 0\} \in \mathbb{R}$, on a $X = a$ presque sûrement. ■

Remarque 3.4.5 On peut généraliser le résultat précédent comme suit : Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et X une variable aléatoire $\sigma(\bigcup_{n \geq m} \sigma(X_n))$ -mesurable pour tout $m \in \mathbb{N}$. Alors X est presque sûrement constant.

Exemple 3.4.6 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes. Alors, les événements

$$\left\{ \sum_{n \in \mathbb{N}} X_n \text{ converge} \right\} \quad \text{et} \quad \left\{ \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n X_k \text{ existe} \right\}$$

sont de probabilité 0 ou 1. Remarquons que les variables aléatoires $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}$ ne sont pas indépendantes mais qu'on peut travailler avec les σ -algèbres indépendantes $\mathcal{F}_n = \sigma(X_n)$, $n \in \mathbb{N}$.

3.5 Indépendance et espérance

Le théorème de Fubini permet de gérer l'intégrale d'un produit de variables aléatoires indépendantes.

Proposition 3.5.1 Une famille X_i , $i \in I$, de variables aléatoires est indépendante si et seulement si pour tout ensemble fini d'indices $J \subset I$ et tout ensemble d'applications mesurables $h_j : \mathbb{R} \rightarrow [0, +\infty[$, $j \in J$, on a

$$\mathbb{E} \left[\prod_{j \in J} h_j(X_j) \right] = \prod_{j \in J} \mathbb{E} [h_j(X_j)].$$

Démonstration : Si les variables aléatoires X_i $i \in I$, sont indépendantes et si $J = \{j_1, \dots, j_p\}$, il est clair que les variables aléatoires X_{j_1}, \dots, X_{j_p} le sont également. On utilise alors le LOTUS et le théorème de Fubini-Tonelli pour écrire

$$\begin{aligned} \mathbb{E} \left[\prod_{k=1}^p h_{j_k}(X_{j_k}) \right] &= \int_{\mathbb{R}^p} \prod_{k=1}^p h_{j_k}(x_{j_k}) d\mathbb{P}_{X_{j_1}}(x_{j_1}) \dots d\mathbb{P}_{X_{j_p}}(x_{j_p}) \\ &= \prod_{k=1}^p \int_{\mathbb{R}} h_{j_k}(x_{j_k}) d\mathbb{P}_{X_{j_k}}(x_{j_k}) \\ &= \prod_{k=1}^p \mathbb{E}[h_{j_k}(X_{j_k})]. \end{aligned}$$

Pour la réciproque, il suffit de prendre pour applications $h_j = \mathbf{1}_{B_j}$ et d'appliquer la Proposition 3.3.2 puisque $\mathbb{E}[h_j(X_j)] = \mathbb{P}(X_j \in B_j)$. ■

Remarque 3.5.2 Si les variables aléatoires X_1, \dots, X_n sont indépendantes et si les applications $h_j : \mathbb{R} \rightarrow \mathbb{R}$ satisfont $h_j(X_j) \in L^1$ pour tout $j \in \{1, \dots, n\}$, alors l'utilisation du théorème de

Tonelli-Fubini comme précédemment permet également d'écrire

$$\mathbb{E} \left[\prod_{j=1}^n h_j(X_j) \right] = \prod_{j=1}^n \mathbb{E} [h_j(X_j)].$$

Pour obtenir la réciproque du théorème précédent, il suffit en fait d'avoir l'égalité pour toutes les fonctions h_j du type $h_j = \mathbb{1}_{B_j}$. Souvent, on demande donc que l'égalité des espérances soit valide pour toute fonction mesurable bornée h_j .

Dans le cas particulier où $h_j = \text{id}$, le résultat précédent montre que le produit de variables aléatoires indépendantes de L^1 reste dans L^1 et que l'espérance du produit est égal au produit des espérances. Notons que cette dernière propriété ne garantit pas l'indépendance. De plus, ces deux propriétés sont fausses si on enlève l'hypothèse d'indépendance.

Corollaire 3.5.3 Si $X_1, \dots, X_n \in L^1$ sont des variables aléatoires indépendantes, alors $X_1 \dots X_n \in L^1$ et

$$\mathbb{E}[X_1 \dots X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n].$$

Corollaire 3.5.4 Si $X_1, \dots, X_n \in L^2$ sont des variables aléatoires indépendantes, alors

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n].$$

Par conséquent, on a l'inégalité de Bienaymé-Tchebychev

$$\mathbb{P} \left(\left| \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right| \geq a \right) \leq \frac{1}{a^2} \sum_{i=1}^n \text{Var}[X_i]$$

pour tout $a > 0$.

Démonstration : Il suffit de le montrer pour deux variables aléatoires X et Y de L^2 (et ensuite de procéder de proche en proche). On peut également supposer que les variables aléatoires X et Y sont centrées puisque la variance est indépendante par translation. On a alors, en utilisant le Corollaire 3.5.3,

$$\begin{aligned} \text{Var}[X + Y] = \mathbb{E}[(X + Y)^2] &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] \\ &= \mathbb{E}[X^2] + \mathbb{E}[Y^2] = \text{Var}[X] + \text{Var}[Y]. \end{aligned}$$

■

Le résultat suivant est immédiat.

Corollaire 3.5.5 Si $X, Y \in L^2$ sont des variables aléatoires indépendantes, alors elles sont non-corrélées.

Remarque 3.5.6 La réciproque de ce résultat est fautive : la non-corrélation n'implique pas l'indépendance. La corrélation donne juste une information concernant la dépendance *linéaire* entre les variables. Par exemple, si $X \sim \mathcal{N}(0, 1)$ et si $Y = X^2$, alors $\mathbb{E}[X] = 0$, $\mathbb{E}[Y] = 1$ et $\mathbb{E}[XY] = \mathbb{E}[X^3] = 0$, d'où $\text{Cov}[X, Y] = 0$. Or X et Y ne sont clairement pas indépendants.

Remarque 3.5.7 Comme $\text{Var}[X] = \|X\|_{L^2}^2$ lorsque X est centré, le résultat précédent nous dit que si X et Y sont des variables aléatoires de L^2 centrées et indépendantes, alors elles sont orthogonales dans l'espace de Hilbert L^2 . Ainsi, dans L^2 , pour des variables aléatoires *centrées*, on a

$$X \perp\!\!\!\perp Y \implies X \perp Y.$$

Terminons cette section par une remarque très utile sur la fonction caractéristique.

Corollaire 3.5.8 Une famille quelconque de variables aléatoires réelles (X_1, \dots, X_n) est indépendante si et seulement si

$$\varphi^{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \varphi^{X_1}(t_1) \cdots \varphi^{X_n}(t_n)$$

pour tout $(t_1, \dots, t_n) \in \mathbb{R}^n$.

Démonstration : Si les variables aléatoires X_1, \dots, X_n sont indépendantes, le résultat est une application de la Proposition 3.5.1 avec les fonctions h_j définies par $h_j(t) = e^{itx_j}$. La réciproque s'obtient comme conséquence du Corollaire 2.6.5 et de la Proposition 3.3.5. ■

3.6 Sommes de variables aléatoires indépendantes

Les sommes de variables aléatoires indépendantes jouent un rôle important en théorie des probabilités. Nous y reviendrons dans le chapitre 5. Dans cette section, nous mentionnons quelques résultats concernant la somme de deux variables aléatoires indépendantes. Afin de pouvoir travailler dans un contexte général, il faut tout d'abord introduire le produit de convolution de mesures. Ensuite, nous étudierons ce qui se passe dans les cas continus et discrets.

Définition 3.6.1 Soient μ et ν deux mesures de probabilité sur $(\mathbb{R}, \mathcal{B})$. On définit le *produit de convolution* $\mu * \nu$ de μ et ν comme étant la mesure image de $\mu \times \nu$ par l'application $+$: $(x, y) \mapsto x + y$.

Le théorème de transfert nous permet d'écrire que

$$\int g d\mu * \nu = \int g(x + y) d\mu \times \nu(x, y)$$

pour toute application $g : \mathbb{R} \rightarrow \mathbb{R}$ mesurable positive ou intégrable.

Proposition 3.6.2 Soient X et Y deux variables aléatoires indépendantes. Alors

- la loi de $X + Y$ est donnée par $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$,
- la fonction de répartition de $X + Y$ est donnée par

$$F_{X+Y}(t) = \int_{\mathbb{R}} F_X(t - y) d\mathbb{P}_Y(y) = \int_{\mathbb{R}} F_Y(t - x) d\mathbb{P}_X(x),$$

- si X est continu de densité f_X , alors $X + Y$ est continu de densité

$$f_{X+Y} = \int f_X(t - y) d\mathbb{P}_Y(y).$$

Démonstration : Puisque X et Y sont indépendants, on a $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \times \mathbb{P}_Y$ et donc

$$\int g(x+y)d\mathbb{P}_{(X,Y)}(x,y) = \int g(x+y)d\mathbb{P}_X \times \mathbb{P}_Y(x,y) = \int g d\mathbb{P}_X * \mathbb{P}_Y$$

pour toute fonction g mesurable et positive. En particulier, pour $g = \mathbb{1}_B$ avec $B \in \mathcal{B}$, on a

$$\mathbb{P}(X+Y \in B) = \int \mathbb{1}_B(x+y)d\mathbb{P}_{(X,Y)}(x,y) = \int \mathbb{1}_B d\mathbb{P}_X * \mathbb{P}_Y = \mathbb{P}_X * \mathbb{P}_Y(B),$$

ce qui montre que la loi de $X+Y$ est bien donnée par $\mathbb{P}_X * \mathbb{P}_Y$. De plus, pour t fixé, si on prend $B = \{z \in \mathbb{R} : z \leq t\}$, il vient

$$\begin{aligned} \mathbb{P}(X+Y \leq t) &= \int \mathbb{1}_B(x+y)d\mathbb{P}_{(X,Y)}(x,y) = \int \mathbb{1}_B(x+y)d\mathbb{P}_X \times \mathbb{P}_Y(x,y) \\ &= \int \left(\int \mathbb{1}_{\{w:w \leq t-y\}}(x)d\mathbb{P}_X(x) \right) d\mathbb{P}_Y(y) \\ &= \int F_X(t-y)d\mathbb{P}_Y(y), \end{aligned}$$

ce qui donne le résultat pour la fonction de répartition. Enfin, si X est continu, on a

$$\begin{aligned} \int g(x+y)d\mathbb{P}_{(X,Y)}(x,y) &= \int \int g(x+y)d\mathbb{P}_X(x)d\mathbb{P}_Y(y) \\ &= \int \int g(x+y)f_X(x)dx d\mathbb{P}_Y(y) \\ &= \int \int g(t)f_X(t-y)dt d\mathbb{P}_Y(y) \\ &= \int g(t) \left(\int f_X(t-y)d\mathbb{P}_Y(y) \right) dt \end{aligned}$$

par Tonelli-Fubini. ■

Remarque 3.6.3 Remarquons qu'il suffit donc que l'un des deux vecteurs ait une densité pour que la somme en ait une.

Remarque 3.6.4 On peut ré-écrire chacune des intégrales ci-dessus comme une espérance, donc $F_{X+Y}(t) = \mathbb{E}[F_Y(t-X)] = \mathbb{E}[F_X(t-Y)]$ et $f_{X+Y}(t) = \mathbb{E}[f_X(t-Y)]$.

Remarque 3.6.5 Le produit de convolution vérifie un certain nombre de propriétés algébriques qui se justifient facilement grâce à la description en terme de variables aléatoires :

- $\mathbb{P} * \delta_0 = \mathbb{P}$ puisque $X+0 = X$.
- Commutativité : $\mathbb{P} * \mathbb{P}' = \mathbb{P}' * \mathbb{P}$ puisque $X+Y = Y+X$.
- Associativité : $(\mathbb{P} * \mathbb{P}') * \mathbb{P}'' = \mathbb{P} * (\mathbb{P}' * \mathbb{P}'')$ puisque $(X+Y)+Z = X+(Y+Z)$.
- Distributivité : $\mathbb{P} * (\alpha\mathbb{P}' + (1-\alpha)\mathbb{P}'') = \alpha(\mathbb{P} * \mathbb{P}') + (1-\alpha)(\mathbb{P} * \mathbb{P}'')$ pour tout $\alpha \in [0,1]$. En effet on utilise les propriétés de l'espérance pour montrer que si g est mesurable et bornée

$$\int g d(\mathbb{P} * (\alpha\mathbb{P}' + (1-\alpha)\mathbb{P}'')) = \int g d(\alpha(\mathbb{P} * \mathbb{P}') + (1-\alpha)(\mathbb{P} * \mathbb{P}'')).$$

Le résultat suivant est une conséquence directe de la proposition précédente. Nous le redémontrons pour en avoir une preuve indépendante.

Proposition 3.6.6 Soient X et Y deux variables aléatoires continues. Si X et Y sont indépendantes, alors $X + Y$ est une variable aléatoire continue dont la densité est donnée par $f_{X+Y} = f_X * f_Y$ où

$$f_X * f_Y(x) = \int_{\mathbb{R}} f_X(x-t)f_Y(t)dt = \int_{\mathbb{R}} f_X(t)f_Y(x-t)dt.$$

Démonstration : Considérons l'application $h : \mathbb{R}^2 \rightarrow \mathbb{R} : (x, y) \mapsto x + y$. Pour tout $B \in \mathcal{B}$, l'indépendance des variables aléatoires X et Y permet d'écrire

$$\begin{aligned} \mathbb{P}(X + Y \in B) &= \mathbb{E}[\mathbf{1}_{h^{-1}(B)}(X, Y)] \\ &= \iint \mathbf{1}_{h^{-1}(B)} f_{(X,Y)}(x, y) dx dy \\ &= \iint \mathbf{1}_{h^{-1}(B)} f_X(x) f_Y(y) dx dy \\ &= \iint_{\{(x,y):x+y \in B\}} f_X(x) f_Y(y) dx dy. \end{aligned}$$

En effectuant le changement de variables $(x, y) \mapsto (t, s) = (x + y, y)$ de jacobien égal à 1, on trouve

$$\mathbb{P}(X + Y \in B) = \int_B \left(\int_{\mathbb{R}} f_X(t-s) f_Y(s) ds \right) dt = \int_B f_X * f_Y(t) dt.$$

■

Remarque 3.6.7 La fonction $f = f_X * f_Y$ s'appelle le *produit de convolution* de f_X et f_Y .

Proposition 3.6.8 Si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ sont indépendantes, alors

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Démonstration : Supposons dans un premier temps que $\mu_1 = \mu_2 = 0$. Alors, la variable aléatoire continue $X + Y$ admet comme fonction de densité la fonction

$$\begin{aligned} f(x) = f_X * f_Y(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} e^{-\frac{(x-t)^2}{2\sigma_1^2}} e^{-\frac{t^2}{2\sigma_2^2}} dt \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma_1^2\sigma_2^2}(\sigma_2^2(x-t)^2 + \sigma_1^2 t^2)} dt. \end{aligned}$$

En posant $\sigma^2 = \sigma_1^2 + \sigma_2^2$ et $\tau = \frac{\sigma_2^2}{\sigma^2}x$, on vérifie que

$$\sigma_2^2(x-t)^2 + \sigma_1^2 t^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma^2} + \sigma^2(t-\tau)^2$$

d'où

$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2}{2\sigma^2}} \int_{\mathbb{R}} e^{-\frac{\sigma_1^2\sigma_2^2}{2\sigma_1^2\sigma_2^2}(t-\tau)^2} dt = \frac{1}{2\pi\sigma} e^{-\frac{x^2}{2\sigma^2}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}} dt = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

en posant $u = \frac{\sigma}{\sigma_1\sigma_2}(t-\tau)$. Pour le cas général, puisqu'on sait que $X - \mu_1 \sim \mathcal{N}(0, \sigma_1^2)$ et que $Y - \mu_2 \sim \mathcal{N}(0, \sigma_2^2)$, ce qui précède implique que $X + Y - \mu_1 - \mu_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$. Cela suffit pour obtenir $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. ■

Nous verrons à la fin de cette section un résultat permettant d'obtenir cette propriété de stabilité de manière bien plus efficace.

Dans le cas discret, on obtient que la fonction de masse de la somme de variables aléatoires indépendantes est égale au produit de convolution discret des fonctions de masse. En effet, si X et Y sont deux variables aléatoires discrètes indépendantes, alors $X + Y$ est une variable aléatoire discrète telle que

$$\mathbb{P}(X + Y = k) = \sum_t \mathbb{P}(X = k - t) \mathbb{P}(Y = t).$$

Proposition 3.6.9 *Si $X \sim \text{Poi}(\lambda)$ et $Y \sim \text{Poi}(\mu)$ sont indépendantes, alors*

$$X + Y \sim \text{Poi}(\lambda + \mu).$$

Démonstration : Clairement, la variable aléatoire $X + Y$ prend ses valeurs dans \mathbb{N} . De plus, pour tout $k \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \sum_{j=0}^k \mathbb{P}(X = k - j) \mathbb{P}(Y = j) \\ &= \sum_{j=0}^k e^{-\lambda} \frac{\lambda^{k-j}}{(k-j)!} e^{-\mu} \frac{\mu^j}{j!} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^k C_k^j \lambda^{k-j} \mu^j \\ &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k, \end{aligned}$$

ce qui suffit. ■

Proposition 3.6.10 *Soient X_1, \dots, X_n des variables aléatoires indépendantes qui suivent une loi de Bernoulli de paramètre p . Alors, la variable aléatoire $X = X_1 + \dots + X_n$ suit une loi binomiale $\text{Bin}(n, p)$.*

Démonstration : Comme chaque variable aléatoire X_j prend ses valeurs dans $\{0, 1\}$, la variable aléatoire X est à valeurs dans $\{0, \dots, n\}$. Pour tout $k \in \{0, \dots, n\}$, on a

$$\mathbb{P}(X = k) = \sum_{(x_1, \dots, x_n) \in \{0, 1\}^n : x_1 + \dots + x_n = k} \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n).$$

Remarquons que si $(x_1, \dots, x_n) \in \{0, 1\}^n$ est tel que $x_1 + \dots + x_n = k$, alors

$$\mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) = p^k (1 - p)^{n-k}.$$

De plus, $\#\{(x_1, \dots, x_n) \in \{0, 1\}^n : x_1 + \dots + x_n = k\} = C_n^k$ et donc

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

■

Corollaire 3.6.11

- Si $X \sim \text{Bin}(n, p)$, alors $\text{Var}[X] = np(1 - p)$.
- Si $X \sim \text{Bin}(n_1, p)$ et $Y \sim \text{Bin}(n_2, p)$ sont indépendantes, alors $X + Y \sim \text{Bin}(n_1 + n_2, p)$.

Toutes les lois qui ont la propriété de “stabilité par sommation indépendante” sont appelées des lois *stables*. Terminons cette section par un résultat général et très pratique.

Proposition 3.6.12 Soient X_1, \dots, X_n des variables aléatoires indépendantes. Alors, on a

$$\varphi^{X_1 + \dots + X_n} = \varphi^{X_1} \dots \varphi^{X_n}.$$

Démonstration : On a

$$\varphi^{X_1 + \dots + X_n}(t) = \mathbb{E}[e^{it(X_1 + \dots + X_n)}] = \mathbb{E}[e^{itX_1} \dots e^{itX_n}] = \mathbb{E}[e^{itX_1}] \dots \mathbb{E}[e^{itX_n}]$$

par la Proposition 3.5.3. ■

Revenons à la somme de deux lois normales indépendantes $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. On sait que $\varphi^X(t) = e^{it\mu_1 - \sigma_1^2 t^2 / 2}$ et $\varphi^Y(t) = e^{it\mu_2 - \sigma_2^2 t^2 / 2}$. En utilisant le résultat précédent, il vient

$$\varphi^{X+Y}(t) = \varphi^X(t)\varphi^Y(t) = e^{it\mu_1 - \sigma_1^2 t^2 / 2} e^{it\mu_2 - \sigma_2^2 t^2 / 2} = e^{it(\mu_1 + \mu_2) - (\sigma_1^2 + \sigma_2^2)t^2 / 2}$$

qui est la fonction caractéristique d’une loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

3.7 Vecteurs aléatoires gaussiens

Dans cette section, nous revenons sur quelques propriétés des vecteurs aléatoires gaussiens précédemment introduits. Commençons par étudier un cas très particulier mais très important.

Proposition 3.7.1 (Exemple fondamental) Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées (i.i.d.) selon une loi $\mathcal{N}(0, 1)$. Alors, le vecteur aléatoire $X = (X_1, \dots, X_n)$ est gaussien. De plus,

- son vecteur moyen est le vecteur nul,
- sa matrice de variance-covariance est la matrice Id ,
- il admet la densité

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2 / 2}.$$

Démonstration : Il suffit de montrer que toute combinaison linéaire des variables X_1, \dots, X_n suit une loi normale. C’est évident car si $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, l’indépendance des variables aléatoires X_1, \dots, X_n et la Proposition 3.6.8 assurent que

$$\sum_{j=1}^n \alpha_j X_j \sim \mathcal{N}\left(0, \sum_{j=1}^n \alpha_j^2\right).$$

Il est évident que son vecteur moyen est nul et la matrice de variance-covariance est l’identité par le Corollaire 3.5.5. Enfin, le dernier point est une conséquence immédiate de la Proposition 3.3.9. ■

Proposition 3.7.2 Si $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_m)$ sont des vecteurs aléatoires gaussiens et si X est indépendant de Y , alors $W = (X, Y)$ est un vecteur aléatoire gaussien.

Démonstration : C'est immédiat car si $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m \in \mathbb{R}$, on sait que $\sum_{j=1}^n \alpha_j X_j$ et $\sum_{k=1}^m \beta_k Y_k$ sont des variables aléatoires indépendantes qui suivent des lois normales. On utilise alors la Proposition 3.6.8 pour conclure. ■

Proposition 3.7.3 Si $X = (X_1, \dots, X_n)^T$ est un vecteur aléatoire gaussien de vecteur moyen μ et de matrice de variance-covariance Σ , alors AX est un vecteur gaussien de vecteur moyen $A\mu$ et de matrice de variance-covariance $A\Sigma A^T$, pour toute matrice réelle A de dimension $m \times n$.

Démonstration : On a $(AX)_j = \sum_{k=1}^n a_{j,k} X_k$ et donc toute combinaison linéaire des composantes de AX peut se réécrire comme combinaison linéaire des composantes de X . Pour la deuxième partie, on utilise la Proposition 2.4.3. ■

Proposition 3.7.4 Si X est un vecteur gaussien de vecteur moyen μ et de matrice de variance-covariance Σ , alors sa fonction caractéristique est donnée par

$$\varphi^X(t) = \exp\left(it^T \mu - \frac{1}{2} t^T \Sigma t\right), \quad t \in \mathbb{R}^n.$$

En particulier, la loi d'un vecteur gaussien est entièrement déterminée par son vecteur moyen μ et sa matrice de variance-covariance Σ . On note $X \sim \mathcal{N}(\mu, \Sigma)$. Si $\mu = 0$, on dit que le vecteur gaussien est centré.

Démonstration : On a $\varphi^X(t) = \mathbb{E}[e^{it^T X}] = \mathbb{E}[e^{i(t_1 X_1 + \dots + t_n X_n)}]$. Or, comme X est un vecteur gaussien, on sait que $t_1 X_1 + \dots + t_n X_n$ suit une loi normale. De plus, on calcule directement que $\mathbb{E}[t_1 X_1 + \dots + t_n X_n] = t_1 \mu_1 + \dots + t_n \mu_n = t^T \mu$ et $\text{Var}[t_1 X_1 + \dots + t_n X_n] = \text{Var}[t^T X] = t^T \Sigma t$. La conclusion s'obtient par calcul de la fonction caractéristique d'une loi univariée normale. ■

Proposition 3.7.5 Un vecteur aléatoire $X = (X_1, \dots, X_n)^T$ est gaussien si et seulement si il existe une matrice carrée réelle D de dimension n et un vecteur $m \in \mathbb{R}^n$ tels que

$$X \stackrel{\mathcal{L}}{=} DN + m,$$

où $N \sim \mathcal{N}(0, \text{Id})$.

Démonstration : Si X est gaussien, alors on pose $m = \mathbb{E}[X]$ et on choisit une matrice D telle que $DD^T = \text{Var}[X]$ (ceci est possible puisque $\text{Var}[X]$ est une matrice réelle symétrique semi-définie positive). Alors $DN + m$ a le même vecteur moyen et la même matrice de variance-covariance que X , ce qui suffit. La réciproque est immédiate. ■

Corollaire 3.7.6 Pour tout $\mu \in \mathbb{R}^n$ et toute matrice carrée Σ de dimension n , symétrique et semi-définie positive, il existe un vecteur gaussien de vecteur moyen μ et de matrice de variance-covariance Σ .

Les vecteurs gaussiens jouissent d'une propriété très importante liée à l'indépendance.

Proposition 3.7.7 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire gaussien. Alors, les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si la matrice de variance-covariance de X est diagonale, i.e. pour tous $i, j \in \{1, \dots, n\}$ tels que $i \neq j$, $\text{Cov}[X_i, X_j] = 0$.

Démonstration : Si les composantes de X sont indépendantes, on sait par le Corollaire 3.5.5 que sa matrice de variance-covariance est diagonale. Réciproquement, si la matrice de variance-covariance de X est diagonale, alors la fonction caractéristique de X est donnée par

$$\varphi^X(t) = \exp\left(it^T \mu - \frac{1}{2}t^T \Sigma t\right) = \exp\left(\sum_{j=1}^n \left(it_j \mathbb{E}[X_j] - \frac{1}{2} \text{Var}[X_j] t_j^2\right)\right) = \prod_{j=1}^n \varphi^{X_j}(t_j).$$

Cela suffit par le Corollaire 3.5.8. ■

Bien entendu, en général une corrélation nulle n'implique en aucune manière l'indépendance. Même dans le cas de vecteurs à marginales gaussiennes, il faut bien veiller à considérer le comportement conjoint des composantes !

Exemple 3.7.8 Considérons $N \sim \mathcal{N}(0, 1)$ et ε une variable aléatoire de loi de Rademacher, c'est-à-dire $\varepsilon \sim \text{Unif}(\{-1, 1\})$. Alors $X_1 = N$ et $X_2 = \varepsilon N$ sont deux variables aléatoires gaussiennes standard, de covariance nulle mais pas indépendantes. Le couple (X_1, X_2) n'est toutefois pas multivarié gaussien.

Définition 3.7.9 Un vecteur gaussien X est *non-dégénéré* si sa matrice de variance-covariance est inversible.

Proposition 3.7.10 Soit $X \sim \mathcal{N}(\mu, \Sigma)$ un vecteur gaussien de dimension n . Alors, X est continu si et seulement si X est non-dégénéré. Dans ce cas, on a

$$f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

pour tout $x \in \mathbb{R}^n$.

Démonstration : Supposons d'abord que X est non-dégénéré. Soit D une matrice telle que $DD^T = \Sigma$. Comme $\det(\Sigma) = (\det D)^2$, on sait que D est inversible. Alors, il est clair que

$$D^{-1}(X - \mu) \sim \mathcal{N}(0, \text{Id}).$$

Il suffit alors d'appliquer la formule du changement de variables donnée à la Proposition 1.5.13. On a

$$\begin{aligned} f_X(x) = f_N(D^{-1}(x - \mu)) \frac{1}{\det D} &= \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(D^{-1}(x - \mu))^T D^{-1}(x - \mu)\right) \\ &= \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T (D^{-1})^T D^{-1}(x - \mu)\right) \end{aligned}$$

où $N \sim \mathcal{N}(0, \text{Id})$. On conclut en utilisant le fait que $(D^{-1})^T D^{-1} = (DD^T)^{-1} = \Sigma^{-1}$.

Supposons à présent que Σ n'est pas inversible. Comme le résultat ne concerne que la loi de X , on peut travailler avec le vecteur aléatoire $DN + \mu$ où $N \sim \mathcal{N}(0, \text{Id})$ et $DD^T = \Sigma$. De plus, on peut sans perte de généralité supposer que $\mu = 0$. On sait que D n'est pas inversible et si $r < n$ est le rang de D , DN est à valeurs dans un sous-espace de \mathbb{R}^n de dimension r . Ainsi, DN ne peut avoir de densité puisque cet espace est de mesure de Lebesgue nulle, alors que pour \mathbb{P}_{DN} , il est de mesure 1. ■

Chapitre 4

Espérances conditionnelles et martingales

4.1 Introduction et motivation

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et A, B deux événements de \mathcal{A} . Si B est de probabilité non nulle, on définit la *probabilité conditionnelle de A sachant B* comme $\mathbb{P}(A | B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$, à savoir la réévaluation de la mesure de A parmi les $\omega \in B$. Nous savons que l'application

$$\mathbb{P}(\cdot | B) : \mathcal{A} \rightarrow [0, 1] : A \mapsto \mathbb{P}(A | B)$$

nous donne une nouvelle mesure de probabilité sur (Ω, \mathcal{A}) ; on peut donc définir des variables aléatoires dans cette nouvelle mesure et considérer toutes les notions étudiées telles que l'espérance ou la variance. En désignant par \mathcal{M} l'ensemble des mesures de probabilité sur (Ω, \mathcal{F}) , nous allons également considérer

$$\mathbb{P}(\cdot | \cdot) : \mathcal{F} \rightarrow \mathcal{M} : B \mapsto \mathbb{P}(\cdot | B)$$

qui assigne à un événement B choisi au hasard selon la loi \mathbb{P} la loi de probabilité conditionnelle étant donné B . On peut également définir des variables aléatoires dans ces lois aléatoires.

De telles *lois conditionnelles* sont omniprésentes dans toutes les applications des probabilités et statistiques. L'objectif de ce chapitre est de les définir et d'en étudier les propriétés les plus importantes. Nous commençons en attaquant les cas "faciles", i.e. ceux dans lesquels une définition intuitive s'offre directement.

4.1.1 Lois conditionnelles discrètes

Soient X, Y deux variables aléatoires discrètes sur $(\Omega, \mathcal{A}, \mathbb{P})$. Notons $\{x_i : i \in I\}$ et les $\{y_j : j \in J\}$ les ensembles des valeurs prises avec une probabilité 1 par X et Y respectivement, avec $I, J \subseteq \mathbb{N}$. Supposons donnée la distribution jointe du vecteur aléatoire bivarié (X, Y) :

$$P_{(X,Y)}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j) = p_{ij}, \quad i \in I, j \in J$$

pour $p_{ij} \in [0, 1]$ tels que $\sum_{i \in I, j \in J} p_{ij} = 1$. Les lois marginales sont alors données par les fonctions de masse

$$P_X(x_i) = \mathbb{P}(X = x_i) = p_{i\bullet} = \sum_{j \in J} p_{ij}, \quad i \in I,$$
$$P_Y(y_j) = \mathbb{P}(Y = y_j) = p_{\bullet j} = \sum_{i \in I} p_{ij}, \quad j \in J.$$

Pour chaque $i \in I$, la distribution de Y conditionnellement à l'événement $\{X = x_i\}$ est donnée par $\mathbb{P}(Y = y_j | X = x_i) = \mathbb{P}(X = x_i, Y = y_j) / \mathbb{P}(X = x_i) = p_{ij} / p_{i\bullet}$. Ceci nous définit, pour chaque $i \in I$, une variable aléatoire discrète $Y | X = x_i$ de fonction de masse

$$P_{Y|X=x_i}(y_j) = \mathbb{P}(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i\bullet}}, \quad j \in J.$$

Comme à toute distribution discrète univariée, on y associe une espérance

$$\mathbb{E}[h(Y) | X = x_i] = \sum_{j \in J} h(y_j) \frac{p_{ij}}{p_{i\bullet}},$$

pour autant que la somme soit absolument convergente et que h soit mesurable. On appelle cette espérance une "espérance conditionnelle". En particulier ceci nous définit une variance conditionnelle :

$$\begin{aligned} \text{Var}[Y | X = x_i] &= \sum_{j \in J} (y_j - \mathbb{E}[Y | X = x_i])^2 \frac{p_{ij}}{p_{i\bullet}} \\ &= \sum_{j \in J} (y_j)^2 \frac{p_{ij}}{p_{i\bullet}} - (\mathbb{E}[Y | X = x_i])^2 \\ &= \mathbb{E}[Y^2 | X = x_i] - (\mathbb{E}[Y | X = x_i])^2, \end{aligned}$$

pour autant que les sommes soient finies. A nouveau, ces quantités sont définies pour chaque $i \in I$.

Remarque 4.1.1 Bien entendu on peut échanger les rôles de Y et de X ci-dessus pour définir pour chaque $j \in J$ la variable aléatoire discrète $X | Y = y_j$ de fonction de masse $P_{X|Y=y_j}$ sur $\{x_i, i \in I\}$, ainsi que les espérances $\mathbb{E}[X | Y = y_j]$ et variances $\text{Var}[X | Y = y_j]$ conditionnelles correspondantes.

Exemple 4.1.2 Considérons l'expérience de l'Exemple 1.5.8 qui consiste à lancer deux dés équilibrés (distinguables), d'espace fondamental $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}$ équipé de la σ -algèbre des parties $\mathcal{F} = \mathcal{P}(\Omega)$, avec X la somme des résultats de chaque dé et Y leur différence en valeur absolue. La distribution jointe est donnée dans la table 4.1. Il est facile d'obtenir les distributions conditionnelles de Y sachant les diverses valeurs de X . Les espérances et variances correspondantes sont alors faciles à déterminer. Le tout est résumé dans la table 4.2. On pourrait également faire le travail pour calculer les lois de X étant données les différentes valeurs de Y .

Si h est une fonction à valeurs réelles telle que $\mathbb{E}[h(Y) | X = x_i]$ est définie pour chaque x_i , la construction ci-dessus mène à la nouvelle fonction à valeurs réelles

$$x_i \mapsto \mathbb{E}[h(Y) | X = x_i].$$

En rendant à X son caractère aléatoire, on introduit alors une nouvelle variable aléatoire discrète : la variable aléatoire *moyenne conditionnelle* $\mathbb{E}[h(Y) | X]$ qui, lorsque $X = x_i$ (c'est-à-dire lorsque $\omega \in \Omega$ est tel que $X(\omega) = x_i$), prend la valeur $\mathbb{E}[h(Y) | X = x_i]$. Plus précisément, comme fonction de ω , la variable aléatoire $\mathbb{E}[h(Y) | X]$ est définie par

$$\omega \mapsto \mathbb{E}[h(Y) | X](\omega) = \mathbb{E}[h(Y) | X(\omega)].$$

	2	3	4	5	6	7	8	9	10	11	12			
0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$	$\frac{6}{36}$	$\mathbb{E}[Y]$ $= \frac{35}{18}$ ≈ 1.94	$\text{Var}[Y]$ $= \frac{665}{324}$ ≈ 2.05
1		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{10}{36}$		
2			$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$			$\frac{8}{36}$		
3				$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$				$\frac{6}{36}$		
4					$\frac{2}{36}$		$\frac{2}{36}$					$\frac{4}{36}$		
5						$\frac{2}{36}$						$\frac{2}{36}$		
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1		
$\mathbb{E}[X] = 7$														
$\text{Var}[X] = \frac{35}{6} \approx 5.83$														

TABLE 4.1 – Loi jointe et lois marginales de la somme et la différence des résultats du lancer de deux dés distinguables.

	2	3	4	5	6	7	8	9	10	11	12
0	1		$\frac{1}{3}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{3}$		1
1		1		$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$		1	
2			$\frac{2}{3}$		$\frac{2}{5}$		$\frac{2}{5}$		$\frac{2}{3}$		
3				$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$			
4					$\frac{2}{5}$		$\frac{2}{5}$				
5						$\frac{1}{3}$					
	1	1	1	1	1	1	1	1	1	1	1
$\mathbb{E}[Y X = x_i]$	0	1	1.33	2	2.4	3	2.4	2	1.33	1	0
$\text{Var}[Y X = x_i]$	0	0	0.89	1	2.24	2.67	2.24	1	0.89	0	0

TABLE 4.2 – Lois conditionnelles de $Y | X = x_i$ pour les différentes valeurs de x_i , ainsi que les espérances et variances correspondantes.

Remarquons que la loi de la variable $Z = \mathbb{E}[h(Y) | X]$ est donnée par

$$\mathbb{P}(Z = \mathbb{E}[h(Y) | X = x_i]) = \mathbb{P}(X = x_i) \text{ pour tout } i \in I.$$

Par extension, on définit également la variable aléatoire *variance conditionnelle* $\text{Var}[Y|X]$ qui, lorsque $X = x_i$, prend la valeur $\text{Var}[Y|X = x_i]$:

$$\omega \mapsto \text{Var}[Y | X](\omega) = \text{Var}[Y | X(\omega)] = \mathbb{E}[Y^2 | X(\omega)] - (\mathbb{E}[Y | X(\omega)])^2.$$

Exemple 4.1.3 A la suite de l'exemple précédent, les deux dernières lignes de la table nous définissent les variables $\mathbb{E}[Y | X]$ et $\text{Var}[Y | X]$ qui prennent les différentes valeurs $\mathbb{E}[Y | X = k]$ et $\text{Var}[Y | X = k]$, $k = 2, \dots, 12$ avec probabilité $\mathbb{P}[X = k]$, $k = 2, \dots, 12$.

Théorème 4.1.4 (Loi des espérances totales) Soit X, Y des variables aléatoires discrètes et h une application mesurable. Si $h(Y) \in L^1$, alors

$$\mathbb{E}[\mathbb{E}[h(Y)|X]] = \mathbb{E}[h(Y)]. \tag{4.1}$$

Démonstration : Par définition, on a

$$\begin{aligned} \mathbb{E}[\mathbb{E}[h(Y)|X]] &= \sum_{i \in I} \mathbb{E}[h(Y)|X = x_i] p_{i\bullet} \\ &= \sum_{i \in I} \left(\sum_{j \in J} h(y_j) \frac{p_{ij}}{p_{i\bullet}} \right) p_{i\bullet} \\ &= \sum_{i \in I} \sum_{j \in J} h(y_j) p_{ij} \\ &= \sum_{j \in J} h(y_j) \left(\sum_{i \in I} p_{ij} \right) \\ &= \sum_{j \in J} h(y_j) p_{\bullet j} = \mathbb{E}[h(Y)] \end{aligned}$$

où l'on peut changer l'ordre de sommation vu que les différentes séries convergent absolument par hypothèse. ■

Exemple 4.1.5 On peut vérifier ce résultat sur l'Exemple 4.1.2 : on calcule que l'espérance de la variable aléatoire $\mathbb{E}[Y | X]$ est donnée par

$$\mathbb{E}[\mathbb{E}[Y | X]] = 0\mathbb{P}(X = 2) + 1\mathbb{P}(X = 3) + 1.33\mathbb{P}(X = 4) + \dots + 1\mathbb{P}(X = 11) + 0\mathbb{P}(X = 12),$$

ce qui donne bien $\mathbb{E}[Y] = \frac{35}{18}$.

Remarque 4.1.6 La loi des probabilités totales "habituelle" s'obtient en corollaire du Théorème 4.1.4 : si on prend la fonction $\mathbb{1}_A$ pour un ensemble borélien A , on applique l'identité pour déduire

$$\mathbb{P}(Y \in A) = \mathbb{E}[\mathbb{1}_A(Y)] = \mathbb{E}[\mathbb{E}[\mathbb{1}_A(Y) | X]] = \sum_{i \in I} \mathbb{P}(Y \in A | X = x_i) \mathbb{P}(X = x_i).$$

Notons qu'il est en général faux que $\mathbb{E}[\text{Var}[Y|X]] = \text{Var}[Y]$. On a en revanche le résultat suivant.

Théorème 4.1.7 (Décomposition de la variance) Soient X, Y des variables aléatoires discrètes. Si $Y \in L^2$, alors

$$\text{Var}[Y] = \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]].$$

Démonstration : Puisque $\text{Var}[Y|X] = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$, on a successivement

$$\begin{aligned} \mathbb{E}[\text{Var}[Y|X]] &= \mathbb{E}[\mathbb{E}[Y^2|X]] - \mathbb{E}[(\mathbb{E}[Y|X])^2] \\ &= \mathbb{E}[Y^2] - \mathbb{E}[(\mathbb{E}[Y|X])^2] \\ &= \text{Var}[Y] + (\mathbb{E}[Y])^2 - \mathbb{E}[(\mathbb{E}[Y|X])^2] \\ &= \text{Var}[Y] + (\mathbb{E}[\mathbb{E}[Y|X]])^2 - \mathbb{E}[(\mathbb{E}[Y|X])^2] \\ &= \text{Var}[Y] - \text{Var}[\mathbb{E}[Y|X]] \end{aligned}$$

où l'on a appliqué deux fois la loi des espérances totales. ■

Remarque 4.1.8 L'identité du théorème précédent, qui est vraie de façon totalement générale dès qu'on a la loi des espérances totales, donne une décomposition de la variance qui s'avère très utile dans une variété de domaines plus appliqués (économie, finance, ...).

4.1.2 Lois conditionnelles continues

Soit (X, Y) un vecteur aléatoire bivarié continu de densité jointe f et de densités marginales f_X et f_Y respectivement. On a alors

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx.$$

Par analogie au cas discret, on adopte la définition suivante.

Définition 4.1.9 Soit (X, Y) un vecteur aléatoire bivarié continu de densité jointe f et de densités marginales f_X et f_Y , respectivement. Les *densités conditionnelles* de $Y|X = x$ et de $X|Y = y$ sont définies sur \mathbb{R} par

$$f_{Y|X=x}(y) = \begin{cases} \frac{f(x, y)}{f_X(x)} & \text{si } f_X(x) \neq 0 \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f_{X|Y=y}(x) = \begin{cases} \frac{f(x, y)}{f_Y(y)} & \text{si } f_Y(y) \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

Remarque 4.1.10 Nous remarquons que, contrairement au cas discret, la définition de densité conditionnelle continue ne découle pas de la définition de probabilité conditionnelle, étant donné que les différents événements sont maintenant de probabilité nulle. La définition choisie est toutefois la seule raisonnable.

On vérifie aisément que, pour chaque x tel que $f_X(x) > 0$, la fonction $y \mapsto f_{Y|X=x}(y)$ est la densité d'une variable aléatoire; nous la noterons $Y|X = x$. Comme pour toute distribution continue univariée, on peut définir une espérance

$$\mathbb{E}[h(Y)|X = x] = \int_{\mathbb{R}} h(y) f_{Y|X=x}(y) dy,$$

pour autant que l'intégrale existe. On appelle cette espérance une "espérance conditionnelle". En particulier, ceci nous définit une "variance conditionnelle" :

$$\begin{aligned} \text{Var}[Y|X = x] &= \int_{\mathbb{R}} (y - \mathbb{E}[Y|X = x])^2 f_{Y|X=x}(y) dy \\ &= \int_{\mathbb{R}} y^2 f_{Y|X=x}(y) dy - (\mathbb{E}[Y|X = x])^2 \\ &= \mathbb{E}[Y^2 | X = x] - (\mathbb{E}[Y | X = x])^2. \end{aligned}$$

Remarquons que toutes ces espérances se définissent en inversant les rôles de X et Y , donc on a également la loi de X conditionnellement à $Y = y$, bien sûr.

Exemple 4.1.11 Supposons que (X, Y) admette la fonction de densité

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{sinon.} \end{cases}$$

Ceci donne en particulier (pour chaque $x \in [0, 1]$)

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \frac{3(x + y^2)}{3x + 1} & \text{si } y \in [0, 1] \\ 0 & \text{sinon,} \end{cases}$$

ce qui livre

$$\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy = \int_0^1 y \frac{3(x + y^2)}{3x + 1} dy = \frac{6x + 3}{12x + 4}$$

et

$$\begin{aligned} \text{Var}[Y|X = x] &= \int_{-\infty}^{\infty} y^2 f_{Y|X=x}(y) dy - (\mathbb{E}[Y|X = x])^2 \\ &= \int_0^1 y^2 \frac{3(x + y^2)}{3x + 1} dy - \left(\frac{6x + 3}{12x + 4} \right)^2 \\ &= \frac{60x^2 + 44x + 3}{80(3x + 1)^2}. \end{aligned}$$

Si h est une fonction telle que $\mathbb{E}[h(Y) | X = x]$ est définie pour chaque x , la construction ci-dessus mène à la nouvelle fonction

$$x \mapsto \mathbb{E}[h(Y) | X = x].$$

En rendant à X son caractère aléatoire, on introduit alors une nouvelle variable aléatoire continue : la *moyenne conditionnelle* $\mathbb{E}[h(Y) | X]$ qui, lorsque $X = x$ (i.e. pour tout $\omega \in \Omega$ tel que $X(\omega) = x$), prend la valeur $\mathbb{E}[h(Y) | X = x]$. Plus précisément, comme fonction de ω , la variable aléatoire $\mathbb{E}[h(Y) | X]$ est définie par

$$\omega \mapsto \mathbb{E}[h(Y) | X](\omega) = \mathbb{E}[h(Y) | X(\omega)].$$

Par extension on définit également la v.a. *variance conditionnelle* $\text{Var}[Y|X]$.

Exemple 4.1.12 En reprenant l'exemple précédent, on a

$$\mathbb{E}[Y|X] = \frac{6X + 3}{12X + 4} \text{ et } \text{Var}[Y | X] = \frac{60X^2 + 44X + 3}{80(3X + 1)^2}.$$

Alors, comme dans le cas discret, on a le résultat suivant.

Théorème 4.1.13 (Loi des espérances totales) Soient X, Y des variables aléatoires continues. Si $Y \in L^2$, alors

- $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$,
- $\text{Var}[Y] = \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]]$.

Etant donné le premier point, la preuve du second s'applique tant au cas continu qu'au cas discret car on n'y fait pas usage de la nature de l'espérance considérée. Formellement, la preuve du premier point est identique à celle du cas discret (on remplace seulement les sommes par des intégrales et on utilise Fubini pour le changement d'ordre d'intégration) et est laissée en exercice.

Exemple 4.1.14 A titre d'illustration, dans le cas de l'exemple précédent, on a

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y|X]] &= \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_X(x) dx \\ &= \int_0^1 \left(\frac{6x + 3}{12x + 4} \right) \left(\frac{2}{5} (3x + 1) \right) dx \\ &= \int_0^1 \frac{1}{10} (6x + 3) dx = \frac{3}{5}, \end{aligned}$$

ce qui coïncide bien avec $\mathbb{E}[Y]$ précédemment calculé.

Nous concluons cette section en remarquant qu'on obtient, comme corollaire immédiat des différentes définitions, l'équivalent continu de la règle de Bayes et de la loi des probabilités totales.

Corollaire 4.1.15 Soient X, Y des variables aléatoires de loi jointe continue.

- Règle de Bayes :

$$f_{Y|X=x}(y) = \frac{f_Y(y)}{f_X(x)} f_{X|Y=y}(x) \tag{4.2}$$

en tout point (x, y) tel que $f_X(x) \neq 0$ et $f_Y(y) \neq 0$;
- Loi des probabilités totales :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y=y}(x) f_Y(y) dy. \tag{4.3}$$

4.1.3 Autres exemples

Evidemment, rien n'impose que toutes les composantes d'un vecteur soient de même nature. Nous verrons plus tard une définition qui s'adapte à tous les cas. Nous traitons néanmoins quelques exemples, histoire de se faire la main.

Exemple 4.1.16 Soit (X, Y) un vecteur avec $X \sim \text{Exp}(1)$ un temps d'attente (en heures) et $Y = \mathbb{1}_{\{X \geq 1\}}$ qui marque si l'on a dû attendre plus d'une heure ou non. Alors les lois marginales sont faciles à déterminer :

$$X \sim \text{Exp}(1) \quad (\text{par définition}) \quad \text{et} \quad Y \sim \text{Bern}(p)$$

avec $p = \mathbb{P}(X \geq 1) = \int_1^\infty e^{-x} dx = 1/e$. La loi jointe s'obtient grâce à

$$\mathbb{P}(X \leq x, Y = 0) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x} & \text{si } 0 \leq x < 1 \\ 1 - e^{-1} & \text{si } x \geq 1 \end{cases} \quad \text{et} \quad \mathbb{P}(X \leq x, Y = 1) = \begin{cases} 0 & \text{si } x < 1 \\ e^{-1} - e^{-x} & \text{si } x \geq 1 \end{cases}.$$

On déduit la densité jointe

$$f_{(X,Y)}(x, 0) = e^{-x} \mathbb{1}_{[0,1[}(x) \quad \text{et} \quad f_{(X,Y)}(x, 1) = e^{-x} \mathbb{1}_{[1,+\infty[}(x).$$

On obtient les densités conditionnelles

$$f_{X|Y=0}(x) = \frac{e^{-x}}{1 - e^{-1}} \mathbb{1}_{[0,1[}(x) \quad \text{et} \quad f_{X|Y=1}(x) = \frac{e^{-x}}{e^{-1}} \mathbb{1}_{[1,+\infty[}(x)$$

de même que

$$\mathbb{P}(Y = 0 | X = x) = \mathbb{1}_{[0,1[}(x) \quad \text{et} \quad \mathbb{P}(Y = 1 | X = x) = \mathbb{1}_{[1,+\infty[}(x).$$

On calcule

$$\mathbb{E}[X | Y = 0] = \frac{e - 2}{e - 1} \approx 0.41 \quad \text{et} \quad \mathbb{E}[X | Y = 1] = 2$$

ainsi que

$$\mathbb{E}[Y | X = x] = \mathbb{1}_{[1,+\infty[}(x).$$

On remarquera que $\mathbb{E}[X | Y]$ est une variable aléatoire discrète (binaire) ; on remarquera également que $\mathbb{E}[Y | X] = Y$.

Exemple 4.1.17 Considérons une variable aléatoire X de loi de Poisson de paramètre Λ , avec Λ lui-même aléatoire de loi exponentielle de moyenne $r > 0$. Alors, la variable aléatoire $X | \Lambda = \lambda$ est Poisson de paramètre λ ; la loi de X est donc une loi "hiérarchique" dont le paramètre est lui-même aléatoire. En utilisant (une version intuitive – justifiée dans la section suivante – de) la loi des probabilités totales, on calcule

$$\mathbb{E}[X] = \int_0^\infty \mathbb{E}[X | \Lambda = \lambda] f^\Lambda(\lambda) d\lambda = \int_0^\infty \lambda \frac{1}{\lambda} e^{-\lambda/r} d\lambda = r$$

ou encore

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | \Lambda]] = \mathbb{E}[\Lambda] = r,$$

ce qui montre donc que le résultat est en fait valable quelle que soit la loi de Λ .

Exemple 4.1.18 Considérons X_1, X_2, X_3 des variables aléatoires indépendantes de loi Bernoulli de paramètre $\frac{1}{2}$. Définissons $S_3 = X_1 + X_2 + X_3$. Alors (X_1, S_3) a une loi jointe non triviale ; $S_3 | X_1 = x$ a une loi (et donc une moyenne et une variance) qui dépend de x

- si $x = 0$ alors $S_3 | X_1 = 0$ est nécessairement égal à $X_2 + X_3$
- si $x = 1$, alors $S_3 | X_1 = 1$ est égal à $1 + X_2 + X_3$.

Par conséquent, on vient d'obtenir que $\mathbb{E}[S_3 | X_1 = 0] = 1$ et $\mathbb{E}[S_3 | X_1 = 1] = 2$. Remarquons donc qu'on a $\mathbb{E}[S_3 | X_1] = X_1 + 1$ et $\text{Var}[S_3 | X_1] = \text{Var}[X_2 + X_3]$.

4.1.4 Espérance et projection

Les sections précédentes nous permettent de définir l'espérance conditionnelle au cas par cas de façon intuitive. Il existe toutefois une autre façon intuitive de définir cette quantité, en se rappelant du Théorème 2.2.11 qui montre que si $X \in L^2$, alors

$$\mathbb{E}[X] = \operatorname{argmin}_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2], \quad (4.4)$$

ce qui signifie que la quantité $\mathbb{E}[X]$ s'interprète donc comme notre meilleure approximation (dans L^2) de la quantité aléatoire X quand on ne dispose d'aucune information supplémentaire sur l'expérience aléatoire qui génère le nombre X . Ce raisonnement nous encourage à définir l'espérance conditionnelle comme la fonction de Y approchant au mieux X dans L^2 , i.e.

$$\mathbb{E}[X | Y] = \operatorname{argmin}_g \mathbb{E}[(X - g(Y))^2] \quad (4.5)$$

où le minimum est pris sur toutes les fonctions mesurables g pour lesquelles $g(Y) \in L^2$. En exprimant les choses en termes de la géométrie de l'espace des variables aléatoires de carré intégrable, c'est la *projection orthogonale dans L^2 de X sur le sous-espace des variables aléatoires $\sigma(Y)$ -mesurables* qui va atteindre le minimum dans (4.5). En effet, si $X = g_0(Y) + Y^\perp$ avec Y^\perp tel que $\mathbb{E}[Y^\perp g(Y)] = 0^1$ pour toute fonction mesurable g telle que $g(Y) \in L^2$, alors

$$\begin{aligned} \mathbb{E}[(X - g(Y))^2] &= \mathbb{E}[(X - g_0(Y) + g_0(Y) - g(Y))^2] \\ &= \mathbb{E}[(X - g_0(Y))^2] + 2\mathbb{E}[(X - g_0(Y))(g_0(Y) - g(Y))] + \mathbb{E}[(g_0(Y) - g(Y))^2] \\ &= \mathbb{E}[(X - g_0(Y))^2] + 2\mathbb{E}[Y^\perp(g_0(Y) - g(Y))] + \mathbb{E}[(g_0(Y) - g(Y))^2] \\ &= \mathbb{E}[(X - g_0(Y))^2] + \mathbb{E}[(g_0(Y) - g(Y))^2] \end{aligned}$$

d'où l'on tire que la projection orthogonale de X sur l'ensemble des variables aléatoires Y mesurables "fait le job".

4.2 L'espérance conditionnelle

Dans cette section, nous définissons l'espérance conditionnelle d'une variable aléatoire de L^1 par rapport à une σ -algèbre \mathcal{G} . Intuitivement, cette espérance conditionnelle est la variable aléatoire \mathcal{G} -mesurable qui est "la plus proche" de la variable aléatoire donnée. Pour cela, nous vérifierons que la définition adoptée donne dans le cas L^2 la projection orthogonale de l'égalité 4.5. Nous montrerons également que le cas général proposé ici permet de récupérer les définitions données dans les sous-sections 4.1.1 et 4.1.3.

4.2.1 σ -algèbre et information

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit $\mathcal{G} \subset \mathcal{F}$ une autre σ -algèbre. Ici et dans la suite nous allons considérer l'analogie " σ -algèbre = information" pour introduire une notion beaucoup plus générale d'espérance conditionnelle à une σ -algèbre plutôt qu'à une simple variable aléatoire.

Exemple 4.2.1 Soient $\Omega = \{a, b, c, d, e\}$ et les variables

1. C'est le produit scalaire entre Y^\perp et $g(Y)$ dans L^2 .

	a	b	c	d	e
X	1	1	1	2	2
Y	1	2	2	3	4
Z	1	2	2	2	3

Alors $\sigma(X) = \sigma\{\{a, b, c\}, \{d, e\}\}$, $\sigma(Y) = \sigma\{\{a\}, \{b, c\}, \{d\}, \{e\}\}$ et $\sigma(Z) = \sigma\{\{a\}, \{b, c, d\}, \{e\}\}$. En se focalisant sur X , on lit plusieurs choses dans les observations ci-dessus :

- tous les événements de $\sigma(X)$ sont de la forme $X \in B$ pour un $B \in \mathcal{B}$. Un événement $A \in \sigma(X)$ est donc une union d'événements de la forme $X = x$; la connaissance de $\omega \in A$ est donc une connaissance sur le comportement de X .
- $\sigma(X) \subseteq \sigma(Y)$, ce qui s'interprète en disant qu'il suffit de connaître Y pour savoir la valeur de X .
- $\sigma(X) \not\subseteq \sigma(Z)$: il ne suffit pas de connaître Z pour connaître X (si $Z = 2$, X peut valoir 1 ou 2).

Définition 4.2.2 Soient X, Y deux variables aléatoires et \mathcal{G} une σ -algèbre. On dit que

- X est \mathcal{G} -mesurable si $\sigma(X) \subseteq \mathcal{G}$, i.e. $X^{-1}(B) \in \mathcal{G}$ pour tout $B \in \mathcal{B}$,
- X est Y -mesurable si X est $\sigma(Y)$ -mesurable,
- X est indépendante de \mathcal{G} (et on écrit $X \perp \mathcal{G}$) lorsque $\sigma(X) \perp \mathcal{G}$.

Exemple 4.2.3 Dans les notations de l'exemple précédent, X est $\sigma(Y)$ -mesurable, Y n'est pas $\sigma(X)$ -mesurable et X n'est pas $\sigma(Z)$ -mesurable.

Le lemme suivant montre que les variables aléatoires Y -mesurables sont les fonctions boréliennes de Y .

Lemme 4.2.4 (de Doob) Soit $Y : \Omega \rightarrow \mathbb{R}$ et soit $X : \Omega \rightarrow \mathbb{R}$. Alors X est Y -mesurable si et seulement s'il existe une application $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ Borel-mesurable telle que $X = h(Y)$.

Démonstration : Supposons que $X = h(Y)$ où $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ est Borel-mesurable. Alors, pour tout ensemble borélien B de $\overline{\mathbb{R}}$, on a

$$X^{-1}(B) = \{X \in B\} = \{h(Y) \in B\} = \{Y \in h^{-1}(B)\} = Y^{-1}(h^{-1}(B)) \in \sigma(Y)$$

puisque $h^{-1}(B) \in \mathcal{B}$ et $\sigma(Y) = \{Y^{-1}(A) : A \in \mathcal{B}\}$.

Pour montrer la réciproque, on commence par étudier le cas où X est une variable aléatoire positive simple Y -mesurable, c'est-à-dire telle que

$$X = \sum_{k=1}^N a_k \mathbb{1}_{A_k}$$

où $N \in \mathbb{N}_0$, $a_k \geq 0$ et $A_k \in \sigma(Y)$ pour tout $k \in \{1, \dots, N\}$. Comme $A_k \in \sigma(Y)$, il existe $B_k \in \mathcal{B}$ tel que $A_k = Y^{-1}(B_k)$. Alors, on a $X = h(Y)$ où

$$h = \sum_{k=1}^N a_k \mathbb{1}_{B_k}$$

est une application Borel-mesurable.

Si X est une variable aléatoire positive Y -mesurable, alors il existe une suite croissante $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires positives simples Y -mesurables qui converge ponctuellement vers X . Pour chaque $n \in \mathbb{N}$, il existe une application positive h_n Borel-mesurable telle que $X_n = h_n(Y)$. On a donc que la suite de fonctions $(h_n(Y))_{n \in \mathbb{N}}$ converge en tout point vers X . En particulier, pour tout $\omega \in \Omega$, la limite $\lim_{n \rightarrow +\infty} h_n(Y(\omega))$ existe. On pose alors

$$h = \limsup_{n \rightarrow +\infty} h_n.$$

Alors h est Borel-mesurable et $X = h(Y)$.

Enfin, dans le cas général, on écrit $X = X^+ - X^-$. Par le second cas, il existe des fonctions h^+, h^- Borel-mesurables telles que $X^+ = h^+(Y)$ et $X^- = h^-(Y)$. Il suffit alors de poser $h = h^+ - h^-$. ■

4.2.2 Espérance conditionnelle à une σ -algèbre

En guise d'illustration, supposons que X et Y sont des variables aléatoires discrètes avec $X \in L^1$. Notons $\{x_i, i \in I\}$ et les $\{y_j, j \in J\}$ les ensembles des valeurs prises avec une probabilité 1 par X et Y respectivement, avec $I, J \subseteq \mathbb{N}$, et considérons la variable aléatoire $Z := \mathbb{E}[X | Y]$ qui prend la valeur $z_j = \mathbb{E}[X | Y = y_j]$ lorsque $Y = y_j$ pour $j \in J$.

On remarque que

- $Z = \mathbb{E}[X | Y]$ est constante sur les ensembles $\{Y = y_j\}$ et on peut donc écrire

$$Z = \sum_{j \in J} \mathbb{E}[X | Y = y_j] \mathbf{1}_{\{Y = y_j\}};$$

en particulier, Z est Y -mesurable ;

- pour tout $j_0 \in J$, on a

$$\begin{aligned} \mathbb{E}[Z \mathbf{1}_{\{Y = y_{j_0}\}}] &= \mathbb{E} \left[\sum_{j \in J} \mathbb{E}[X | Y = y_j] \mathbf{1}_{\{Y = y_j\}} \mathbf{1}_{\{Y = y_{j_0}\}} \right] \\ &= \mathbb{E}[X | Y = y_{j_0}] \mathbb{P}(Y = y_{j_0}) \\ &= \sum_{i \in I} x_i \mathbb{P}(X = x_i | Y = y_{j_0}) \mathbb{P}(Y = y_{j_0}) \\ &= \sum_{i \in I} x_i \mathbb{P}(X = x_i, Y = y_{j_0}) \\ &= \mathbb{E} \left[X \mathbf{1}_{\{Y = y_{j_0}\}} \right]; \end{aligned}$$

on en tire que pour tout $G \in \sigma(Y)$ et par linéarité de l'espérance, on a

$$\mathbb{E}[Z \mathbf{1}_G] = \mathbb{E}[X \mathbf{1}_G]$$

étant donné que tout $G \in \sigma(Y)$ s'écrit comme union d'événements de la forme $\{Y = y_j\}$.

Etonnamment, ces deux propriétés sont exactement ce qu'il faut demander pour obtenir la définition correcte d'une espérance conditionnelle générale.

Théorème 4.2.5 (Définition de l'espérance conditionnelle) Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, X une variable aléatoire de L^1 et $\mathcal{G} \subset \mathcal{F}$ une σ -algèbre. Alors, il existe une variable aléatoire \mathcal{G} -mesurable Z telle que

$$\mathbb{E}[Z\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_G] \quad \text{pour tout } G \in \mathcal{G}. \quad (4.6)$$

Si de plus \tilde{Z} est une autre variable aléatoire \mathcal{G} -mesurable intégrable satisfaisant (4.6), alors $Z = \tilde{Z}$ \mathbb{P} -presque sûrement. On appelle Z (une version de) l'espérance conditionnelle de X étant donnée \mathcal{G} et on la note

$$\mathbb{E}[X | \mathcal{G}].$$

La condition (4.6) se réécrit donc

$$\int_G Z d\mathbb{P} = \int_G X d\mathbb{P} \quad \text{pour tout } G \in \mathcal{G}.$$

Démonstration : Commençons par démontrer l'unicité \mathbb{P} -presque sûre de l'espérance conditionnelle. Si \tilde{Z} est une autre variable aléatoire \mathcal{G} -mesurable qui vérifie (4.6), alors on a évidemment $\mathbb{E}[Z\mathbf{1}_G] = \mathbb{E}[\tilde{Z}\mathbf{1}_G]$ pour tout $G \in \mathcal{G}$. Montrons que $\mathbb{P}(Z > \tilde{Z}) = 0$. Par le même argument, on aura $\mathbb{P}(\tilde{Z} > Z) = 0$ d'où $\mathbb{P}(Z \neq \tilde{Z}) = 0$.

Soit $A_n = \{Z - \tilde{Z} \geq 1/n\}$. Clairement $A_n \in \mathcal{G}$ car Z et \tilde{Z} sont \mathcal{G} -mesurables, donc (4.6) s'applique et

$$0 = \mathbb{E}[Z\mathbf{1}_{A_n}] - \mathbb{E}[\tilde{Z}\mathbf{1}_{A_n}] = \mathbb{E}[(Z - \tilde{Z})\mathbf{1}_{A_n}] \geq \frac{1}{n}\mathbb{P}(A_n)$$

ce qui nous apprend que $\mathbb{P}(A_n) = 0$. La suite d'événements A_n étant croissante, on en tire que

$$\mathbb{P}(Z > \tilde{Z}) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}_0} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = 0,$$

ce qui suffit.

Pour l'existence, supposons dans un premier temps que $X \geq 0$ et définissons pour tout $G \in \mathcal{G}$

$$\nu(G) = \mathbb{E}[X\mathbf{1}_G] = \int_G X d\mathbb{P}.$$

Alors ν est une mesure sur (Ω, \mathcal{G}) qui est absolument continue par rapport à \mathbb{P} . Le théorème de Radon-Nikodym A.7.6 donne l'existence d'une application \mathcal{G} -mesurable $Z : \Omega \rightarrow [0, +\infty[$ \mathbb{P} -presque sûrement unique telle que $\nu(G) = \int_G Z d\mathbb{P}$ pour tout $G \in \mathcal{G}$. On a donc

$$\mathbb{E}[X\mathbf{1}_G] = \mathbb{E}[Z\mathbf{1}_G]$$

pour tout $G \in \mathcal{G}$ et donc $Z = \mathbb{E}[X | \mathcal{G}]$ convient. Si X n'est pas positive, on prend $X = X^+ - X^-$ et on définit $Z = \mathbb{E}[X^+ | \mathcal{G}] - \mathbb{E}[X^- | \mathcal{G}]$. ■

Démonstration alternative : On peut également prouver le théorème 4.2.5 sans faire appel au théorème de Radon-Nikodym. Cette preuve est intéressante à développer, même si nous n'irons pas tout à fait jusqu'au bout du raisonnement car nous ne prouverons l'existence que sous l'hypothèse plus forte $X \in L^2$ (alors que l'énoncé ne prévoit l'existence que d'un moment

d'ordre 1) et faisons appel à certains résultats de la théorie des espaces de Hilbert que nous ne re-prouvons pas ici (ils sont développés à la Section II.6 de [1] ou encore à la section 6.9 de [11]).

Considérons l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ de toutes les variables aléatoires de carré intégrable et munissons-le du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$. On peut montrer que l'espace est maintenant un espace de Hilbert (lorsqu'on a pris le quotient par les variables aléatoires égales presque sûrement) et que $L^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sous-espace fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. On peut donc parler de la projection orthogonale $P^\perp(X)$ de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$ qui est une fonction \mathcal{G} -mesurable vérifiant

$$\langle X - P^\perp(X), U \rangle = \mathbb{E}[(X - P^\perp(X))U] = 0 \text{ pour tout } U \in L^2(\Omega, \mathcal{G}, \mathbb{P}).$$

En appliquant cette dernière relation à $U = \mathbb{1}_G$ avec $G \in \mathcal{G}$, il suit que $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[P^\perp(X)\mathbb{1}_G]$ pour tout $G \in \mathcal{G}$, donc la projection orthogonale $P^\perp(X)$ – qui est unique à des ensembles de \mathbb{P} -mesure nulle près – remplit les conditions demandées par la définition. On prend donc $P^\perp(X)$ comme version de l'espérance conditionnelle.

Dans le cas où X n'est supposée qu'intégrable, on la suppose positive et on définit pour tout $n \in \mathbb{N}$ la variable aléatoire $X_n = \min\{X, n\}$ qui est bornée donc dans L^2 . On définit alors $\mathbb{E}[X_n | \mathcal{G}]$ à l'aide de l'argument précédent, et il reste à montrer la convergence vers une variable aléatoire satisfaisant les hypothèses. Le cas où X est quelconque se résout en prenant les parties positives et négatives. Nous ne détaillons pas ici. ■

La définition d'espérance conditionnelle à une variable aléatoire se déduit immédiatement.

Définition 4.2.6 (Espérance conditionnelle à une variable aléatoire) Soient X et Y des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ et supposons que $X \in L^1$. L'espérance conditionnelle de X étant donnée Y est la variable aléatoire

$$\mathbb{E}[X | Y] := \mathbb{E}[X | \sigma(Y)].$$

Plus généralement, si Y_1, \dots, Y_n sont des variables aléatoires, on définit

$$\mathbb{E}[X | Y_1, \dots, Y_n] := \mathbb{E}[X | \sigma(Y_1, \dots, Y_n)].$$

Le lemme de Doob nous offre la caractérisation suivante.

Définition 4.2.7 Si $X \in L^1$, alors $\mathbb{E}[X | Y]$ est l'unique (\mathbb{P} -pp) variable aléatoire $h(Y)$ telle que $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[h(Y)\mathbb{1}_A]$ pour tout $A \in \sigma(Y)$.

4.3 Propriétés de l'espérance conditionnelle

Dans cette section, nous présentons les propriétés de l'espérance conditionnelle. Nous fixons un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une sous- σ -algèbre $\mathcal{G} \subseteq \mathcal{F}$. Comme l'espérance conditionnelle est définie \mathbb{P} -presque partout (i.e. presque sûrement), toutes les relations données sont valides presque sûrement.

Proposition 4.3.1 (Intégrabilité et loi des espérances totales) Si $X \in L^1$, alors

- $\mathbb{E}[X | \mathcal{G}] \in L^1$,
- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}]]$.

Démonstration : Notons $Z = \mathbb{E}[X | \mathcal{G}]$ et remarquons que

$$\mathbb{E}[|Z|] = \mathbb{E}[Z\mathbf{1}_{A^+}] - \mathbb{E}[Z\mathbf{1}_{A^-}]$$

avec $A^+ = \{Z \geq 0\}$ et $A^- = \{Z < 0\}$. Alors, $A^+ \in \mathcal{G}$ et $A^- \in \mathcal{G}$, et donc

$$\begin{aligned} \mathbb{E}[|Z|] &= \mathbb{E}[Z\mathbf{1}_{A^+}] - \mathbb{E}[Z\mathbf{1}_{A^-}] \\ &= \mathbb{E}[X\mathbf{1}_{A^+}] - \mathbb{E}[X\mathbf{1}_{A^-}] \\ &\leq \mathbb{E}[|X|\mathbf{1}_{A^+}] + \mathbb{E}[|X|\mathbf{1}_{A^-}] \\ &= \mathbb{E}[|X|] < +\infty, \end{aligned}$$

où le passage de la première à la deuxième ligne suit directement de (4.6). La deuxième partie suit en prenant $A = \Omega$ dans (4.6). ■

Proposition 4.3.2 (Mesurabilité et indépendance) Soit $X \in L^1$.

- Si X est \mathcal{G} -mesurable alors $\mathbb{E}[X | \mathcal{G}] = X$.
- Si X est indépendante de \mathcal{G} , alors $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X]$.

Démonstration : La première affirmation suit par unicité de l'espérance conditionnelle (et du fait que (4.6) est trivialement satisfait pour $Z = X$). Pour la seconde affirmation, on remarque que la variable aléatoire constante $\mathbb{E}[X]$ est trivialement \mathcal{G} -mesurable. De plus, si $G \in \mathcal{G}$, $\mathbb{E}[\mathbb{E}[X]\mathbf{1}_G] = \mathbb{E}[\mathbf{1}_G]\mathbb{E}[X] = \mathbb{E}[\mathbf{1}_G X]$ par indépendance. ■

Proposition 4.3.3 (Linéarité) Soient $X, Y \in L^1$. Si $a, b \in \mathbb{R}$ alors

$$\mathbb{E}[aX + bY | \mathcal{G}] = a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}].$$

Démonstration : La variable aléatoire $a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}]$ est la somme de deux fonctions \mathcal{G} -mesurables donc également \mathcal{G} -mesurable. De plus, si $G \in \mathcal{G}$, alors

$$\begin{aligned} \mathbb{E}[(a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}])\mathbf{1}_G] &= a\mathbb{E}[\mathbb{E}[X | \mathcal{G}]\mathbf{1}_G] + b\mathbb{E}[\mathbb{E}[Y | \mathcal{G}]\mathbf{1}_G] \\ &= a\mathbb{E}[X\mathbf{1}_G] + b\mathbb{E}[Y\mathbf{1}_G] \\ &= \mathbb{E}[(aX + bY)\mathbf{1}_G] \end{aligned}$$

où la première et la dernière égalité utilisent la linéarité des espérances et la deuxième égalité utilise par deux fois (4.6). Le résultat suit par unicité. ■

Proposition 4.3.4 (Monotonie) Soit $X \in L^1$. Si $X \geq 0$, alors $\mathbb{E}[X | \mathcal{G}] \geq 0$. En particulier,

- si $X \geq Y$, alors $\mathbb{E}[X | \mathcal{G}] \geq \mathbb{E}[Y | \mathcal{G}]$,
- $|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}]$.

Démonstration : Notons $Z = \mathbb{E}[X | \mathcal{G}]$ et définissons $A_n = \{Z \leq -1/n\} \in \mathcal{G}$. Si $\mathbb{P}(Z < 0) > 0$, alors il existe n tel que $\mathbb{P}(A_n) > 0$ puisque $\{Z > 0\} = \bigcup_{n \in \mathbb{N}_0} A_n$. Etant donné que $A_n \in \mathcal{G}$, on a

$$0 \leq \mathbb{E}[X \mathbf{1}_{A_n}] = \mathbb{E}[Z \mathbf{1}_{A_n}] \leq -\frac{1}{n} \mathbb{P}(A_n) < 0,$$

ce qui mène à une contradiction. La deuxième conclusion suit en appliquant la positivité à la variable $(X - Y)$ qui est positive.

Enfin, on a $|X| + X \geq 0$ et $|X| - X \geq 0$ d'où $\mathbb{E}[|X| | \mathcal{G}] \geq -\mathbb{E}[X | \mathcal{G}]$ et $\mathbb{E}[|X| | \mathcal{G}] \geq \mathbb{E}[X | \mathcal{G}]$, ce qui suffit pour conclure. ■

Proposition 4.3.5 (Théorème de la convergence monotone conditionnel) Si $(X_n)_{n \in \mathbb{N}}$ est une suite croissante de variables aléatoires positives de L^1 qui converge \mathbb{P} -p.s. vers $X \in L^1$, alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n | \mathcal{G}] = \mathbb{E}[X | \mathcal{G}]$$

\mathbb{P} -p.s.

Démonstration : Par monotonie de l'espérance conditionnelle, la suite $(\mathbb{E}[X_n | \mathcal{G}])_{n \in \mathbb{N}}$ est croissante et majorée par $\mathbb{E}[X | \mathcal{G}]$. Elle converge donc vers une variable aléatoire Z \mathcal{G} -mesurable. Montrons que $Z = \mathbb{E}[X | \mathcal{G}]$. Si $G \in \mathcal{G}$, on a

$$\begin{aligned} \mathbb{E}[Z \mathbf{1}_G] &= \mathbb{E} \left[\lim_{n \rightarrow +\infty} \mathbb{E}[X_n | \mathcal{G}] \mathbf{1}_G \right] = \lim_{n \rightarrow +\infty} \mathbb{E} [\mathbb{E}[X_n | \mathcal{G}] \mathbf{1}_G] \\ &= \lim_{n \rightarrow +\infty} \mathbb{E} [X_n \mathbf{1}_G] \\ &= \mathbb{E} \left[\lim_{n \rightarrow +\infty} X_n \mathbf{1}_G \right] \\ &= \mathbb{E} [X \mathbf{1}_G] \end{aligned}$$

où on a utilisé deux fois le théorème de la convergence monotone classique et la définition de l'espérance conditionnelle de X_n étant donnée \mathcal{G} . ■

Remarque 4.3.6 L'hypothèse $X \in L^1$ peut être levée en élargissant la définition de l'espérance conditionnelle au cas des variables aléatoires positives. Si X est une variable aléatoire positive, on définit l'espérance conditionnelle de X étant donnée \mathcal{G} par

$$\mathbb{E}[X | \mathcal{G}] := \lim_{n \rightarrow +\infty} \mathbb{E}[\min\{X, n\} | \mathcal{G}].$$

Des arguments semblables à ceux développés dans la preuve du Théorème de la convergence monotone conditionnel permettent de montrer que cette variable aléatoire est l'unique (\mathbb{P} -p.s.) variable aléatoire \mathcal{G} -mesurable telle que

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbf{1}_G] = \mathbb{E}[X \mathbf{1}_G] \quad \text{pour tout } G \in \mathcal{G}.$$

La même remarque s'applique pour le Lemme de Fatou ci-dessous.

Proposition 4.3.7 (Lemme de Fatou conditionnel) Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires positives de L^1 telle que $\liminf_{n \rightarrow +\infty} X_n \in L^1$, alors

$$\mathbb{E}[\liminf_{n \rightarrow +\infty} X_n \mid \mathcal{G}] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n \mid \mathcal{G}]$$

\mathbb{P} -p.s.

Démonstration : En utilisant le Théorème de convergence monotone conditionnel pour la suite $(\inf_{n \geq k} X_n)_{k \in \mathbb{N}}$, on a

$$\begin{aligned} \mathbb{E}[\liminf_{n \rightarrow +\infty} X_n \mid \mathcal{G}] &= \mathbb{E}[\lim_{k \rightarrow +\infty} \inf_{n \geq k} X_n \mid \mathcal{G}] = \lim_{k \rightarrow +\infty} \mathbb{E}[\inf_{n \geq k} X_n \mid \mathcal{G}] \\ &\leq \lim_{k \rightarrow +\infty} \inf_{n \geq k} \mathbb{E}[X_n \mid \mathcal{G}] \\ &= \liminf_{k \rightarrow +\infty} \mathbb{E}[X_n \mid \mathcal{G}] \end{aligned}$$

par monotonie de l'espérance conditionnelle. ■

Proposition 4.3.8 (On sort ce qu'on connaît) Soient $X, Y \in L^1$ tels que $XY \in L^1$. Si Y est \mathcal{G} -mesurable, alors

$$\mathbb{E}[XY \mid \mathcal{G}] = Y \mathbb{E}[X \mid \mathcal{G}].$$

Démonstration : Clairement $Y \mathbb{E}[X \mid \mathcal{G}]$ est \mathcal{G} -mesurable car c'est le produit de deux variables aléatoires \mathcal{G} -mesurables. Il reste à montrer que pour tout $G \in \mathcal{G}$

$$\mathbb{E}[Y \mathbb{E}[X \mid \mathcal{G}] \mathbf{1}_G] = \mathbb{E}[XY \mathbf{1}_G].$$

Commençons par le cas où Y est simple. Dans ce cas, on a

$$Y = \sum_{k=1}^N a_k \mathbf{1}_{G_k}$$

avec $N \in \mathbb{N}_0$, $a_k \in \mathbb{R}$ et $G_k = Y^{-1}(a_k) \in \mathcal{G}$ pour tout $k \in \{1, \dots, N\}$. Alors, pour tout $G \in \mathcal{G}$ fixé, il vient

$$\begin{aligned} \mathbb{E}[Y \mathbb{E}[X \mid \mathcal{G}] \mathbf{1}_G] &= \sum_{k=1}^N a_k \mathbb{E}[\mathbf{1}_{G_k} \mathbb{E}[X \mid \mathcal{G}] \mathbf{1}_G] \\ &= \sum_{k=1}^N a_k \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mathbf{1}_{G \cap G_k}] \\ &= \sum_{k=1}^N a_k \mathbb{E}[X \mathbf{1}_{G \cap G_k}] \\ &= \mathbb{E}\left[\sum_{k=1}^N a_k X \mathbf{1}_{G \cap G_k}\right] \\ &= \mathbb{E}[XY \mathbf{1}_G] \end{aligned}$$

où on a utilisé la définition de $\mathbb{E}[X | \mathcal{G}]$ puisque $G \cap G_k \in \mathcal{G}$.

Si $Y \geq 0$, on considère une suite croissante $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires \mathcal{G} -mesurables simples qui converge vers Y . Par la première partie, on sait que $\mathbb{E}[XY_n | \mathcal{G}] = Y_n \mathbb{E}[X | \mathcal{G}]$ pour tout $n \in \mathbb{N}$. Par le théorème de la convergence monotone conditionnel, on a

$$\mathbb{E}[XY_n | \mathcal{G}] = \mathbb{E}[X^+ Y_n | \mathcal{G}] - \mathbb{E}[X^- Y_n | \mathcal{G}] \longrightarrow \mathbb{E}[X^+ Y | \mathcal{G}] - \mathbb{E}[X^- Y | \mathcal{G}] = \mathbb{E}[XY | \mathcal{G}]$$

si n tend vers l'infini. D'autre part, on a aussi

$$\mathbb{E}[XY_n | \mathcal{G}] = Y_n \mathbb{E}[X | \mathcal{G}] \longrightarrow Y \mathbb{E}[X | \mathcal{G}],$$

ce qui permet de conclure le deuxième cas. Pour le cas général, il suffit d'écrire $Y = Y^+ - Y^-$ et de se ramener au cas précédent. ■

Proposition 4.3.9 (Théorème de la convergence dominée conditionnel) Soit $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires de L^1 qui converge \mathbb{P} -p.s. vers X . S'il existe une variable aléatoire positive $Z \in L^1$ telle que $|X_n| \leq Z$ pour tout $n \in \mathbb{N}$, alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n | \mathcal{G}] = \mathbb{E}[X | \mathcal{G}]$$

\mathbb{P} -p.s.

Démonstration : Par le théorème de la convergence dominée classique, on sait que $X \in L^1$. Posons $Y_n = 2Z - |X - X_n|$. Alors $(Y_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires positives qui converge presque sûrement vers $2Z$. Le Lemme de Fatou conditionnel implique que

$$\mathbb{E}[2Z | \mathcal{G}] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[2Z - |X - X_n| | \mathcal{G}] = \mathbb{E}[2Z | \mathcal{G}] - \limsup_{n \rightarrow +\infty} \mathbb{E}[|X - X_n| | \mathcal{G}]$$

On a donc

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X| | \mathcal{G}] = 0$$

et par monotonie, on obtient

$$\lim_{n \rightarrow +\infty} |\mathbb{E}[X_n | \mathcal{G}] - \mathbb{E}[X | \mathcal{G}]| \leq \lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X| | \mathcal{G}] = 0.$$

Remarque 4.3.10 Comme d'habitude, on peut supposer que les relations $|X_n| \leq Z$ n'ont lieu que presque sûrement. ■

Proposition 4.3.11 (Transitivité – Tower Property) Si $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ sont des σ -algèbres et si $X \in L^1$, alors on a

$$\mathbb{E}[X | \mathcal{H}] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{H}].$$

Démonstration : Posons $Z = \mathbb{E}[X | \mathcal{H}]$ et montrons que $Z = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{H}]$. Par définition de l'espérance conditionnelle, on sait que Z est \mathcal{H} -mesurable. Si $H \in \mathcal{H}$, il reste à montrer que

$$\mathbb{E}[Z\mathbf{1}_H] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}]\mathbf{1}_H].$$

En appliquant la définition de l'espérance conditionnelle et en utilisant le fait que $H \in \mathcal{G}$, on a

$$\mathbb{E}[Z\mathbf{1}_H] = \mathbb{E}[\mathbb{E}[X | \mathcal{H}]\mathbf{1}_H] = \mathbb{E}[X\mathbf{1}_H] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}]\mathbf{1}_H]$$

et la conclusion suit. ■

Proposition 4.3.12 (Inégalité de Jensen conditionnelle) Soit $X \in L^1$ et soit $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe telle que $\Phi(X) \in L^1$. Alors

$$\mathbb{E}[\Phi(X) | \mathcal{G}] \geq \Phi(\mathbb{E}[X | \mathcal{G}]).$$

Démonstration : On reprend la même preuve que l'inégalité de Jensen classique. On sait que pour tout $t \in \mathbb{R}$, il existe un nombre réel $\beta = \beta(t)$ tel que

$$\Phi(y) \geq \Phi(t) + \beta(y - t)$$

pour tout $y \in \mathbb{R}$. Remarquons que comme $\beta(t)$ peut être choisi comme étant la dérivée à gauche (par exemple) de Φ en t , la fonction $\beta : \mathbb{R} \rightarrow \mathbb{R}$ est croissante (puisque Φ est convexe) et donc Borel-mesurable. En particulier, on a

$$\Phi(X) \geq \Phi(\mathbb{E}[X | \mathcal{G}]) + \beta(X - \mathbb{E}[X | \mathcal{G}]) \quad \mathbb{P} - p.s.$$

où $\beta = \beta(\mathbb{E}[X | \mathcal{G}])$ est une variable aléatoire \mathcal{G} -mesurable. En prenant l'espérance conditionnelle de chaque côté et en utilisant la monotonie de celle-ci, on obtient

$$\begin{aligned} \mathbb{E}[\Phi(X) | \mathcal{G}] &\geq \Phi(\mathbb{E}[X | \mathcal{G}]) + \mathbb{E}[\beta(X - \mathbb{E}[X | \mathcal{G}]) | \mathcal{G}] \\ &= \Phi(\mathbb{E}[X | \mathcal{G}]) + \beta \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}]) | \mathcal{G}] \\ &= \Phi(\mathbb{E}[X | \mathcal{G}]) + \beta \mathbb{E}[X | \mathcal{G}] - \beta \mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{G}] \\ &= \Phi(\mathbb{E}[X | \mathcal{G}]) \end{aligned}$$

où le passage de la première à la seconde ligne est dû au Lemme 4.3.8. ■

Remarque 4.3.13 On retrouve en particulier que $|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}]$ pour toute variable aléatoire $X \in L^1$.

Proposition 4.3.14 (Projection orthogonale) Si $X \in L^2$, alors $\mathbb{E}[X | \mathcal{G}]$ est la projection orthogonale de X sur le sous-espace de $L^2(\Omega, \mathcal{G}, \mathbb{P})$ dans l'espace de Hilbert $L^2(\Omega, \mathcal{F}, \mathbb{P})$ équipé du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$. En d'autres termes,

$$\mathbb{E}[X | \mathcal{G}] = \operatorname{argmin} \{ \mathbb{E}[(X - W)^2] \}$$

où le minimum est pris sur l'ensemble des variables aléatoires $W \in L^2$ \mathcal{G} -mesurables.

Démonstration : Commençons par remarquer que si on note $Z = \mathbb{E}[X | \mathcal{G}]$, alors $Z \in L^2$. En effet, par l'inégalité de Jensen conditionnelle, on sait que

$$Z^2 = (\mathbb{E}[X | \mathcal{G}])^2 \leq \mathbb{E}[X^2 | \mathcal{G}] \in L^1.$$

Soit $W \in L^2$ une variable \mathcal{G} -mesurable. Alors, on a

$$\mathbb{E}[(X - W)^2] = \mathbb{E}[(X - Z)^2] + \mathbb{E}[(Z - W)^2] + 2\mathbb{E}[(X - Z)(Z - W)]$$

et de plus,

$$\begin{aligned} \mathbb{E}[(X - Z)(Z - W)] &= \mathbb{E}[\mathbb{E}[(X - Z)(Z - W) | \mathcal{G}]] \\ &= \mathbb{E}[(Z - W)\mathbb{E}[(X - Z) | \mathcal{G}]] \\ &= \mathbb{E}[(Z - W)(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[Z | \mathcal{G}])] \\ &= \mathbb{E}[(Z - W)(\mathbb{E}[X | \mathcal{G}] - Z)] \\ &= 0 \end{aligned}$$

en utilisant la loi des espérances totales et le fait que $Z - W$ et $Z = \mathbb{E}[X | \mathcal{G}]$ sont \mathcal{G} -mesurables. Par conséquent, on déduit que

$$\mathbb{E}[(X - W)^2] = \mathbb{E}[(X - Z)^2] + \mathbb{E}[(Z - W)^2]$$

d'où il ressort que l'approximation est minimale lorsque $W = Z = \mathbb{E}[X | \mathcal{G}]$. Ceci définit la projection orthogonale, comme demandé. ■

Terminons cette section par quelques remarques supplémentaires.

Remarque 4.3.15 Si $X \in L^2$, alors on définit la variance conditionnelle

$$\text{Var}[X | \mathcal{G}] = \mathbb{E}[X^2 | \mathcal{G}] - (\mathbb{E}[X | \mathcal{G}])^2 = \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2 | \mathcal{G}].$$

On a

$$\mathbb{E}[\text{Var}[X | \mathcal{G}]] = \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2] =: \text{MMSE}(X, \mathcal{G})^2.$$

On a aussi

$$\mathbb{E}[\text{Var}[X | \mathcal{G}]] = \text{Var}[X] - \text{Var}[\mathbb{E}[X | \mathcal{G}]].$$

Remarque 4.3.16 Pour tout $A \in \mathcal{F}$ et par analogie avec la relation $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$, on écrit souvent

$$\mathbb{P}(A | \mathcal{G}) := \mathbb{E}[\mathbb{1}_A | \mathcal{G}].$$

Il faut être prudent car l'objet $\mathbb{P}(A | \mathcal{G})$ est à présent une variable aléatoire.

4.4 Calcul d'espérances conditionnelles

4.4.1 Accord avec les définitions intuitives

Montrons maintenant que cette définition permet de récupérer les définitions intuitives d'espérance conditionnelle développées à la Section 4.1. Rappelons que si Y est une variable aléatoire, l'espérance conditionnelle de X par rapport à Y est $\mathbb{E}[X | \sigma(Y)]$. C'est donc une variable aléatoire $\sigma(Y)$ -mesurable et donc par le Lemme de Doob, une fonction borélienne de Y .

2. Minimum mean square error.

Proposition 4.4.1 (Cas discret) Soient X, Y deux variables aléatoires discrètes sur $(\Omega, \mathcal{A}, \mathbb{P})$ et soient $\{x_i : i \in I\}$ et $\{y_j : j \in J\}$ les ensembles des valeurs prises avec une probabilité 1 par X et Y respectivement, avec $I, J \subseteq \mathbb{N}$. Si $X \in L^1$, on pose

$$h(y_j) = \mathbb{E}[X | Y = y_j] = \sum_{i \in I} x_i \mathbb{P}(X = x_i | Y = y_j), \quad \forall j \in J$$

et $h = 0$ en dehors de $\{y_j, j \in J\}$. Alors,

$$h(Y) = \mathbb{E}[X | Y].$$

Démonstration : Comme $X \in L^1$, alors $\mathbb{E}[X | Y = y_j]$ est fini pour tout $j \in J$. De plus, remarquons que

$$h(y) = \sum_{j \in J} \mathbb{1}_{\{y=y_j\}} \mathbb{E}[X | Y = y_j] = \sum_{j \in J} \mathbb{1}_{\{y=y_j\}} \sum_{i \in I} x_i \mathbb{P}(X = x_i | Y = y_j)$$

et donc la variable aléatoire $h(Y)$ est mesurable par rapport à Y grâce au Lemme de Doob. Il reste donc à montrer que $\mathbb{E}[h(Y)\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$ pour tout $A \in \sigma(Y)$. Remarquons d'abord que

$$\mathbb{E}[h(Y)\mathbb{1}_{\{Y=y_j\}}] = h(y_j)\mathbb{P}(Y = y_j) \text{ pour tout } j \in J$$

car $h(Y)$ est constante si $Y = y_j$. Soit $A \in \sigma(Y)$. Alors $A = \cup_{k \in K} \{Y = y_k\}$ pour un choix d'indices $K \subset J$, ce qui implique que $\mathbb{1}_A = \sum_{k \in K} \mathbb{1}_{\{Y=y_k\}}$ et

$$\mathbb{E}[h(Y)\mathbb{1}_A] = \sum_{k \in K} \mathbb{E}[h(Y)\mathbb{1}_{\{Y=y_k\}}] = \sum_{k \in K} \mathbb{E}[X | Y = y_k] \mathbb{P}(Y = y_k).$$

En utilisant la définition de l'espérance, on tire que

$$\begin{aligned} \mathbb{E}[h(Y)\mathbb{1}_A] &= \sum_{k \in K} \sum_{i \in I} x_i \mathbb{P}(X = x_i | Y = y_k) \mathbb{P}(Y = y_k) \\ &= \sum_{k \in K} \sum_{i \in I} x_i \mathbb{P}(X = x_i, Y = y_k) \\ &= \sum_{i \in I} \sum_{k \in K} x_i \mathbb{P}(X = x_i, Y = y_k) \\ &= \sum_{i \in I} x_i \sum_{k \in K} \mathbb{P}(X = x_i, Y = y_k) \\ &= \sum_{i \in I} x_i \mathbb{P}(X = x_i, Y \in \{y_k, k \in K\}) \\ &= \mathbb{E}[X\mathbb{1}_A], \end{aligned}$$

ce qui suffit. ■

Exemple 4.4.2 Soient $X \sim \text{Pois}(\lambda)$ et $Y = 2\lfloor X/2 \rfloor$. On a directement que $\mathbb{E}[Y | X] = Y$ car Y est $\sigma(X)$ -mesurable. Pour calculer $\mathbb{E}[X | Y]$, on remarque que Y ne prend que des valeurs paires et que $Y = 2n$ si et seulement si $X = 2n$ ou $X = 2n + 1$. Par conséquent, on a

$$\begin{aligned}
 \mathbb{E}[X | Y = 2n] &= \sum_{k \in \mathbb{N}} k \mathbb{P}(X = k | Y = 2n) \\
 &= 2n \mathbb{P}(X = 2n | Y = 2n) + (2n + 1) \mathbb{P}(X = 2n + 1 | Y = 2n) \\
 &= 2n \frac{\mathbb{P}(X = 2n, Y = 2n)}{\mathbb{P}(Y = 2n)} + (2n + 1) \frac{\mathbb{P}(X = 2n + 1, Y = 2n)}{\mathbb{P}(Y = 2n)} \\
 &= 2n \frac{\mathbb{P}(X = 2n)}{\mathbb{P}(X = 2n) + \mathbb{P}(X = 2n + 1)} + (2n + 1) \frac{\mathbb{P}(X = 2n + 1)}{\mathbb{P}(X = 2n) + \mathbb{P}(X = 2n + 1)} \\
 &= 2n \frac{\frac{\lambda^{2n} e^{-\lambda}}{(2n)!}}{\frac{\lambda^{2n} e^{-\lambda}}{(2n)!} + \frac{\lambda^{2n+1} e^{-\lambda}}{(2n+1)!}} + (2n + 1) \frac{\frac{\lambda^{2n+1} e^{-\lambda}}{(2n+1)!}}{\frac{\lambda^{2n} e^{-\lambda}}{(2n)!} + \frac{\lambda^{2n+1} e^{-\lambda}}{(2n+1)!}} \\
 &= 2n \frac{1}{1 + \frac{\lambda}{2n+1}} + (2n + 1) \frac{\frac{\lambda}{2n+1}}{1 + \frac{\lambda}{2n+1}} \\
 &= \frac{(2n + 1)(2n + \lambda)}{2n + 1 + \lambda}
 \end{aligned}$$

et donc

$$\mathbb{E}[X | Y] = \frac{(Y + 1)(Y + \lambda)}{Y + 1 + \lambda}.$$

Dans le cas continu, on retrouve également les densités que l'on avait posées par analogie avec le cas discret. Rappelons que

$$f_{X|Y=y}(x) = \begin{cases} \frac{f(x, y)}{f_Y(y)} & \text{si } f_Y(y) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

vu la Définition 4.1.9.

Proposition 4.4.3 (Cas continu) Soit (X, Y) un vecteur aléatoire bivarié continu tel que $X \in L^1$. Définissons

$$h(y) = \mathbb{E}[X | Y = y] = \int_{\mathbb{R}} x f_{X|Y=y}(x) dx.$$

Alors, on a

$$h(Y) = \mathbb{E}[X | Y].$$

Démonstration : On peut réécrire

$$h(y) = \mathbb{1}_{B_0}(y) \int_{\mathbb{R}} x \frac{f(x, y)}{f_Y(y)} dx = \mathbb{1}_{B_0}(y) \frac{1}{\int_{\mathbb{R}} f(x, y) dx} \int_{\mathbb{R}} x f(x, y) dx$$

où $B_0 = \{y \in \mathbb{R} : f_Y(y) \neq 0\}$, avec $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ Borel-mesurable. Par le Théorème de Tonelli-Fubini, on sait que les deux intégrales apparaissant dans la définition de h sont Borel-mesurables,

et donc h l'est également. Par le Lemme de Doob, on en tire que $h(Y)$ est $\sigma(Y)$ -mesurable. De plus, tout $G \in \sigma(Y)$ est de la forme $G = \{Y \in B\}$ pour $B \in \mathcal{B}(\mathbb{R})$. Donc, on a

$$\begin{aligned} \mathbb{E}[h(Y)\mathbf{1}_G] &= \mathbb{E}[h(Y)\mathbf{1}_{\{Y \in B\}}] = \int_B h(y)f_Y(y)dy = \int_B \int_{\mathbb{R}} xf_{X|Y=y}(x)dx f_Y(y)dy \\ &= \int_B \int_{\mathbb{R}} xf(x,y)dxdy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_B(y)xf(x,y)dxdy \\ &= \mathbb{E}[X\mathbf{1}_{\{Y \in B\}}] \\ &= \mathbb{E}[X\mathbf{1}_G], \end{aligned}$$

d'où la conclusion. ■

4.4.2 Cas d'une σ -algèbre discrète

Supposons que $\mathcal{G} = \sigma(\{A_j : j \in J\})$ où $(A_j)_{j \in J}$ est une partition dénombrable de Ω . Alors, on vérifie directement que \mathcal{G} est la collection de toutes les unions dénombrable d'événements A_j (puisque les complémentaires des ensembles A_j sont également des unions dénombrables de tels événements).

Exemple 4.4.4 Supposons que $\mathcal{G} = \sigma(\{A_1, A_2, A_3, A_4\})$ avec $\Omega = A_1 \cup A_2 \cup A_3 \cup A_4$ et $A_i \cap A_j = \emptyset$ pour tout $i \neq j$. Alors, $A_1^c = A_2 \cup A_3 \cup A_4$.

Proposition 4.4.5 *Supposons que $\mathcal{G} = \sigma(\{A_j : j \in J\})$ où $(A_j)_{j \in J}$ est une partition dénombrable de Ω . Si $X \in L^1$, alors*

$$\mathbb{E}[X | \mathcal{G}] = \sum_{j \in J: \mathbb{P}(A_j) \neq 0} \frac{1}{\mathbb{P}(A_j)} \mathbb{E}[X\mathbf{1}_{A_j}] \mathbf{1}_{A_j}.$$

Démonstration : Posons $Z = \sum_{j \in J: \mathbb{P}(A_j) \neq 0} \frac{1}{\mathbb{P}(A_j)} \mathbb{E}[X\mathbf{1}_{A_j}] \mathbf{1}_{A_j}$. Puisque $A_j \in \mathcal{G}$ pour tout $j \in J$, il est clair que Z est \mathcal{G} -mesurable. De plus, pour tout $k \in J$, on a

$$\begin{aligned} \mathbb{E}[Z\mathbf{1}_{A_k}] &= \sum_{j \in J: \mathbb{P}(A_j) \neq 0} \frac{1}{\mathbb{P}(A_j)} \mathbb{E}[X\mathbf{1}_{A_j}] \mathbb{E}[\mathbf{1}_{A_k}\mathbf{1}_{A_j}] \\ &= \begin{cases} \frac{1}{\mathbb{P}(A_k)} \mathbb{E}[X\mathbf{1}_{A_k}] \mathbb{P}(A_k) & \text{si } \mathbb{P}(A_k) \neq 0 \\ 0 & \text{sinon} \end{cases} \\ &= \mathbb{E}[X\mathbf{1}_{A_k}] \end{aligned}$$

puisque si $\mathbb{P}(A_k) = 0$, alors $X\mathbf{1}_{A_k} = 0$ presque sûrement. Par conséquent, si $G = \bigcup_{k \in K} A_k$ avec $K \subseteq J$ est un événement quelconque de \mathcal{G} , on a

$$\mathbb{E}[Z\mathbf{1}_G] = \mathbb{E} \left[Z \sum_{k \in K} \mathbf{1}_{A_k} \right] = \sum_{k \in K} \mathbb{E}[Z\mathbf{1}_{A_k}] = \sum_{k \in K} \mathbb{E}[X\mathbf{1}_{A_k}] = \mathbb{E} \left[X \sum_{k \in K} \mathbf{1}_{A_k} \right] = \mathbb{E}[X\mathbf{1}_G]$$

ce qui permet de conclure. Remarquons que l'échange entre espérance et série peut se faire en utilisant le théorème de la convergence dominée en majorant par $|Z|$ puisque

$$\mathbb{E}[|Z|] \leq \sum_{j \in J: \mathbb{P}(A_j) \neq 0} \frac{1}{\mathbb{P}(A_j)} \mathbb{E}[|X| \mathbf{1}_{A_j}] \mathbb{E}[\mathbf{1}_{A_j}] = \sum_{j \in J: \mathbb{P}(A_j) \neq 0} \mathbb{E}[|X| \mathbf{1}_{A_j}] = \mathbb{E}[|X|] < +\infty.$$

■

Exemple 4.4.6 Si $\mathcal{G} = \{\emptyset, A, A^c, \Omega\} = \sigma(\{A, A^c\})$ avec $\mathbb{P}(A) \in]0, 1[$, alors pour tout $B \in \mathcal{F}$ on calcule directement

$$\begin{aligned} \mathbb{E}[\mathbf{1}_B | \mathcal{G}] &= \frac{1}{\mathbb{P}(A)} \mathbb{E}[\mathbf{1}_B \mathbf{1}_A] \mathbf{1}_A + \frac{1}{\mathbb{P}(A^c)} \mathbb{E}[\mathbf{1}_B \mathbf{1}_{A^c}] \mathbf{1}_{A^c} \\ &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \mathbf{1}_A + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} \mathbf{1}_{A^c} \\ &= \mathbb{P}(B|A) \mathbf{1}_A + \mathbb{P}(B|A^c) \mathbf{1}_{A^c}. \end{aligned}$$

Exemple 4.4.7 (Loi des espérances totales) Si Y est une variable aléatoire discrète à valeurs dans $\{y_j : j \in J\}$, alors $\sigma(Y) = \sigma(\{Y^{-1}(\{y_j\}) : j \in J\})$. Pour tout $X \in L^1$, on a alors

$$\mathbb{E}[X | Y] = \sum_{j \in J: \mathbb{P}(Y=y_j) \neq 0} \frac{1}{\mathbb{P}(Y=y_j)} \mathbb{E}[X \mathbf{1}_{\{Y=y_j\}}] \mathbf{1}_{\{Y=y_j\}}.$$

En particulier, par la loi des espérances totales, on sait que $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]$ et donc

$$\mathbb{E}[X] = \sum_{j \in J: \mathbb{P}(Y=y_j) \neq 0} \frac{1}{\mathbb{P}(Y=y_j)} \mathbb{E}[X \mathbf{1}_{\{Y=y_j\}}] \mathbb{E}[\mathbf{1}_{\{Y=y_j\}}] = \sum_{j \in J} \mathbb{E}[X \mathbf{1}_{\{Y=y_j\}}]$$

puisque $X \mathbf{1}_{\{Y=y_j\}} = 0$ presque sûrement si $\mathbb{P}(Y=y_j) = 0$. Si X est discret à valeurs presque sûrement dans $\{x_i : i \in I\}$, on a

$$\mathbb{E}[X \mathbf{1}_{\{Y=y_j\}}] = \sum_{i \in I} x_i \mathbb{P}(X = x_i, Y = y_j)$$

et donc

$$\begin{aligned} \mathbb{E}[X] &= \sum_{j \in J} \sum_{i \in I} x_i \mathbb{P}(X = x_i, Y = y_j) = \sum_{j \in J} \sum_{i \in I} x_i \mathbb{P}(X = x_i | Y = y_j) \mathbb{P}(Y = y_j) \\ &= \sum_{j \in J} \mathbb{P}(Y = y_j) \mathbb{E}[X | Y = y_j] \end{aligned}$$

et on retrouve le Théorème 4.1.4.

4.4.3 Conditionnement gaussien

Nous concluons la section en nous concentrant sur le cas particulier des vecteurs gaussiens, dont les propriétés simplificatrices rendent les calculs relativement explicites.

Théorème 4.4.8 Si (X, Y_1, \dots, Y_n) est un vecteur gaussien centré, alors il existe des constantes $a_1, \dots, a_n \in \mathbb{R}$ telles que

$$\mathbb{E}[X | Y_1, \dots, Y_n] = a_1 Y_1 + \dots + a_n Y_n.$$

En particulier, $\mathbb{E}[X | Y_1, \dots, Y_n]$ est gaussien.

Démonstration : Soit E le sous-espace vectoriel de L^2 de dimension finie engendré par Y_1, \dots, Y_n . Sur cet espace, on peut considérer le produit scalaire défini par

$$\langle Z_1, Z_2 \rangle = \mathbb{E}[Z_1 Z_2].$$

On sait alors qu'on peut trouver une base orthonormée Z_1, \dots, Z_m de E avec $m \leq n$. Posons

$$T = \sum_{j=1}^m \mathbb{E}[X Z_j] Z_j.$$

En particulier, $T \in E$ et donc T peut s'écrire comme combinaison linéaire des variables aléatoires Y_1, \dots, Y_n . Montrons que $T - X \in E^\perp$ (en fait, T est la projection orthogonale de X sur E). En effet, pour tout $k \in \{1, \dots, m\}$, on a

$$\begin{aligned} \langle (T - X), Z_k \rangle &= \mathbb{E}[(T - X)Z_k] = \mathbb{E} \left[\sum_{j=1}^m \mathbb{E}[X Z_j] Z_j Z_k - X Z_k \right] \\ &= \sum_{j=1}^m \mathbb{E}[X Z_j] \mathbb{E}[Z_j Z_k] - \mathbb{E}[X Z_k] \\ &= \mathbb{E}[X Z_k] - \mathbb{E}[X Z_k] = 0 \end{aligned}$$

où on a utilisé le fait que la base Z_1, \dots, Z_m est orthonormée. En particulier, on a $\mathbb{E}[(T - X)Y_j] = 0$ pour tout $j \in \{1, \dots, n\}$ et donc les variables aléatoires $T - X$ et Y_j sont non-corrélées. Comme le vecteur (X, Y_1, \dots, Y_n) est gaussien, il en est de même pour le vecteur $(T - X, Y_1, \dots, Y_n)$ et on en tire que $T - X$ est indépendant de Y_1, \dots, Y_n .

Montrons que $T = \mathbb{E}[X | Y_1, \dots, Y_n]$. En utilisant les propriétés de l'espérance conditionnelles vis-à-vis de l'indépendance et de la mesurabilité, on trouve

$$\begin{aligned} \mathbb{E}[X | Y_1, \dots, Y_n] &= \mathbb{E}[X - T | Y_1, \dots, Y_n] + \mathbb{E}[T | Y_1, \dots, Y_n] \\ &= \mathbb{E}[X - T] + T = \mathbb{E}[X] - \mathbb{E}[T] + T \\ &= T \end{aligned}$$

puisque X et T sont des variables centrées. ■

Exemple 4.4.9 Soit (X, Y, Z) un vecteur gaussien centré de matrice de covariance

$$\Sigma = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix}.$$

Pour calculer $\mathbb{E}[Y | X, Z]$ nous utilisons ce qui précède pour déduire que $\mathbb{E}[Y | X, Z] = \alpha X + \beta Z$ et donc

$$\begin{aligned} X\mathbb{E}[Y | X, Z] &= \alpha X^2 + \beta XZ \\ Z\mathbb{E}[Y | X, Z] &= \alpha XZ + \beta Z^2 \end{aligned}$$

. En prenant les espérances de chaque côté et en utilisant les règles de calcul, on obtient

$$\begin{aligned} \mathbb{E}[XY] &= \alpha\mathbb{E}[X^2] + \beta\mathbb{E}[XZ] \\ \mathbb{E}[ZY] &= \alpha\mathbb{E}[XZ] + \beta\mathbb{E}[Z^2] \end{aligned}$$

d'où ressort le système

$$\begin{cases} 0 &= \alpha - \beta \\ 3 &= -\alpha + 4\beta. \end{cases}$$

Il vient $\alpha = \beta = 1$ et $\mathbb{E}[Y | X, Z] = X + Z$.

Remarque 4.4.10

- On a donc obtenu que $\mathbb{E}[X | Y_1, \dots, Y_n]$ est une fonction *linéaire* de Y_1, \dots, Y_n alors qu'en toute généralité, on peut seulement dire qu'il s'agit d'une fonction Borel-mesurable.
- Si Y_1, \dots, Y_n ne sont pas linéairement indépendantes, alors la décomposition de $\mathbb{E}[X | Y_1, \dots, Y_n]$ comme combinaison linéaire de Y_1, \dots, Y_n n'est pas unique.
- Si le vecteur gaussien n'est pas centré, on peut s'y ramener en posant $X' = X - \mathbb{E}[X]$ et $Y'_j = Y_j - \mathbb{E}[Y_j]$. Pour le vecteur gaussien (X', Y'_1, \dots, Y'_n) , le résultat précédent implique l'existence de constantes a_1, \dots, a_n telles que

$$\mathbb{E}[X' | Y'_1, \dots, Y'_n] = a_1 Y'_1 + \dots + a_n Y'_n.$$

Par conséquent, puisque $\sigma(Y'_1 + \mathbb{E}[Y_1], \dots, Y'_n + \mathbb{E}[Y_n]) = \sigma(Y'_1, \dots, Y'_n)$, on a

$$\begin{aligned} \mathbb{E}[X | Y_1, \dots, Y_n] &= \mathbb{E}[X' + \mathbb{E}[X] | Y'_1 + \mathbb{E}[Y_1], \dots, Y'_n + \mathbb{E}[Y_n]] \\ &= \mathbb{E}[X' | Y'_1, \dots, Y'_n] + \mathbb{E}[X] \\ &= \sum_{j=1}^n a_j Y'_j + \mathbb{E}[X] \\ &= \sum_{j=1}^n a_j Y_j + \left(\mathbb{E}[X] - \sum_{j=1}^n a_j \mathbb{E}[Y_j] \right) \end{aligned}$$

et donc $\mathbb{E}[X | Y_1, \dots, Y_n]$ est une fonction *affine* de Y_1, \dots, Y_n .

4.5 Martingales

Nous ne présentons ici qu’une très brève introduction à la théorie des martingales. Il s’agit d’une notion fondamentale de la théorie des probabilités qui tire son origine en théorie des jeux et introduit la notion d’*évolution dans le temps*. Au fur et à mesure que le temps passe, nous avons à notre disposition de plus en plus d’information. On est ainsi amené à considérer des sous- σ -algèbres croissantes d’événements et à chaque instant, on aura une variable aléatoire avec de l’information sur elle “dans le passé”. Nous nous contentons ici d’aborder le cas des martingales à *temps discret*, c’est-à-dire pour lesquelles le temps est indexé par les entiers. Cela revient à regarder une évolution à intervalles réguliers. Si on veut modéliser ce qui se passe à tout instant, il faut alors introduire des martingales à *temps continu*, c’est-à-dire indexées par $[0, +\infty[$.

Définition 4.5.1 Soit (Ω, \mathcal{F}) un espace mesurable. Une suite de sous- σ -algèbres $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ de \mathcal{F} est appelée une *filtration* si $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ pour tout $n \in \mathbb{N}_0$.

Intuitivement, \mathcal{F}_n contient les événements qui peuvent survenir jusqu’à l’instant n . Il faut donc comprendre \mathcal{F}_n comme l’information dont on dispose au temps n : plus le temps passe, plus nous disposons d’information. Par exemple, considérons un jeu de hasard qui est répété un grand nombre de fois : \mathcal{F}_n est l’information dont nous disposons après n parties et la suite de sous- σ -algèbres est bien croissante si nous gagnons (ou à tout le moins ne perdons pas) des informations avec le temps.

Définition 4.5.2 Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ une filtration. Une suite de variables aléatoires $(Z_n)_{n \in \mathbb{N}_0}$ est *adaptée à la filtration* $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ si Z_n est \mathcal{F}_n -mesurable pour tout $n \in \mathbb{N}_0$.

On a typiquement $\mathcal{F}_n = \sigma(W_1, \dots, W_n)$ où les W_i sont des variables aléatoires, et la suite $(Z_n)_{n \in \mathbb{N}_0}$ adaptée à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ est de la forme

$$Z_n = f_n(W_1, \dots, W_n)$$

pour une fonction f_n Borel-mesurable.

Introduisons à présent la notion de martingale. Une martingale peut se voir comme la généralisation du jeu équitable (i.e. un pari dont l’espérance de gain est nulle) : $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale pour $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ si quels que soient les résultats des n premières parties, la fortune attendue après le $n + 1^{\text{ème}}$ jeu est exactement ce qu’elle était avant. En d’autres termes, quel que soit l’historique du jeu le gain du joueur est, en moyenne, nul.

Dans la suite, nous considérons un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ fixé.

Définition 4.5.3 Une suite $(Z_n)_{n \in \mathbb{N}_0}$ de variables aléatoires est une *martingale* pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ si

1. $(Z_n)_{n \in \mathbb{N}_0}$ est adapté à $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$,
2. $Z_n \in L^1$ pour tout $n \in \mathbb{N}_0$,
3. $\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = Z_n$ pour tout $n \in \mathbb{N}_0$.

Remarquons que les deux premiers points permettent de donner du sens au troisième.

En prenant l’espérance de part et d’autre de la dernière condition de la Définition 4.5.3, on déduit

$$\mathbb{E}[Z_{n+1}] = \mathbb{E}[Z_n]$$

et donc, en particulier, $\mathbb{E}[Z_n] = \mathbb{E}[Z_1]$ pour tout $n \in \mathbb{N}_0$. On appelle $\mathbb{E}[Z_1]$ la *moyenne de la martingale*. Notons qu'on a également

$$\mathbb{E}[Z_{n+2} | \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[Z_{n+2} | \mathcal{F}_{n+1}] | \mathcal{F}_n] = \mathbb{E}[Z_{n+1} | \mathcal{F}_n] = Z_n$$

et donc, en répétant,

$$\mathbb{E}[Z_m | \mathcal{F}_n] = Z_n \quad \forall m \geq n.$$

Les martingales sont donc en particulier des suites de variables aléatoires constantes en espérance.

Remarque 4.5.4 On peut également définir la notion de martingale sans filtration préalable. A toute suite de variables aléatoires $Z = (Z_n)_{n \in \mathbb{N}_0}$, on peut associer sa *filtration naturelle*, donnée par

$$\mathcal{F}_n^Z = \sigma(Z_1, \dots, Z_n).$$

On dit que $(Z_n)_{n \in \mathbb{N}_0}$ est une *martingale* si elle est une martingale pour sa filtration naturelle, c'est-à-dire si

1. $Z_n \in L^1$ pour tout $n \in \mathbb{N}_0$,
2. $\mathbb{E}[Z_{n+1} | Z_1, \dots, Z_n] = Z_n$ pour tout $n \in \mathbb{N}_0$.

Soit $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$ une filtration. Notons que si $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale pour $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$, alors c'est nécessairement une martingale pour \mathcal{F}_n^Z : en effet, on a

$$\mathcal{F}_n^Z \subseteq \mathcal{G}_n$$

puisque $(Z_n)_{n \in \mathbb{N}_0}$ est adapté à $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$ (la filtration naturelle est la plus petite filtration adaptée). De plus, en utilisant la propriété d'emboîtement de l'espérance conditionnelle, l'hypothèse que $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale pour $(\mathcal{G}_n)_{n \in \mathbb{N}_0}$ et la mesurabilité de Z_n par rapport à \mathcal{F}_n^Z , on a

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n^Z] = \mathbb{E}[\mathbb{E}[Z_{n+1} | \mathcal{G}_n] | \mathcal{F}_n^Z] = \mathbb{E}[Z_n | \mathcal{F}_n^Z] = Z_n.$$

Présentons à présent quelques exemples simples de martingales.

Exemple 4.5.5

1. Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires indépendantes de L^1 d'espérance nulle. Alors

$$Z_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N}_0$$

est une martingale.

2. Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires indépendantes de L^1 d'espérance 1. Alors

$$Z_n = \prod_{i=1}^n X_i, \quad n \in \mathbb{N}_0$$

est une martingale.

3. Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de L^2 d'espérance nulle et de variance σ^2 . Alors

$$Z_n = \left(\sum_{i=1}^n X_i \right)^2 - n\sigma^2, \quad n \in \mathbb{N}_0$$

est une martingale pour la filtration naturelle de $(X_n)_{n \in \mathbb{N}_0}$.

4. Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires indépendantes de L^1 d'espérance nulle et $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ leur filtration naturelle. Pour chaque $n \in \mathbb{N}_0$, considérons une variable aléatoire B_n bornée et mesurable pour \mathcal{F}_{n-1} ; on peut voir B_n comme un pari sur X_n . La fortune totale à l'instant n est $W_0 = 0$ et

$$W_n = \sum_{j=1}^n B_j X_j.$$

Alors, la suite $(W_n)_{n \in \mathbb{N}_0}$ est une martingale pour la filtration naturelle de $(X_n)_{n \in \mathbb{N}_0}$.

Exemple 4.5.6 (Martingale de Doob) Soit Y une variable aléatoire telle que $\mathbb{E}[|Y|] < +\infty$. Soit \mathcal{F}_n une filtration et

$$Z_n = \mathbb{E}[Y | \mathcal{F}_n].$$

Alors $(Z_n)_{n \in \mathbb{N}}$ est une martingale dans la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. En effet, les deux premiers points sont immédiats et pour le troisième, il suffit de remarquer que

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[Y | \mathcal{F}_{n+1}] | \mathcal{F}_n] = \mathbb{E}[Y | \mathcal{F}_n] = Z_n$$

car $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. On dit que $(Z_n)_{n \in \mathbb{N}}$ est une martingale de Doob.

A la place de considérer un jeu équitable, on peut aussi s'intéresser à des jeux qui sont favorables au joueur, ou au contraire défavorables.

Définition 4.5.7 Une suite de variables aléatoires $(Z_n)_{n \in \mathbb{N}_0}$ forme une *sous-martingale* pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ si

1. $(Z_n)_{n \in \mathbb{N}_0}$ est adapté à $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$,
2. $Z_n \in L^1$ pour tout $n \in \mathbb{N}_0$,
3. $\mathbb{E}[Z_{n+1} | \mathcal{F}_n] \geq Z_n$.

Si l'inégalité \geq dans le dernier point est remplacée par \leq , on parle de *sur-martingale*.

Bien entendu, $(Z_n)_{n \in \mathbb{N}_0}$ est une sous-martingale si et seulement si $(-Z_n)_{n \in \mathbb{N}_0}$ est une sur-martingale. De plus, un processus adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ est une martingale si et seulement si c'est à la fois un sous-martingale et une sur-martingale. Pour cette raison, nous nous intéresserons parfois uniquement à des propriétés sur les sous-martingales, les propriétés correspondantes des sur-martingales et des martingales pourront alors s'en déduire aisément.

Une sous-martingale (resp. sur-martingale) est donc un processus croissant (resp. décroissant) en moyenne conditionnelle et donc en moyenne. En effet, il suit de la tower property des espérances conditionnelles que

$$\mathbb{E}[Z_{n+1}] \geq \mathbb{E}[Z_n]$$

pour une sous-martingale, et de manière plus générale,

$$\mathbb{E}[Z_m | \mathcal{F}_n] \geq Z_n, \quad \forall m \geq n.$$

De même, si $(Z_n)_{n \in \mathbb{N}_0}$ est une sur-martingale, alors

$$\mathbb{E}[Z_m | \mathcal{F}_n] \leq Z_n, \quad \forall m \geq n.$$

Comme pour les martingales, on peut parler de sous-martingale (resp. de sur-martingale) sans spécifier une filtration : il s'agit alors d'une sous-martingale (resp. une sur-martingale) pour la filtration naturelle.

Exemple 4.5.8 (Marche aléatoire) Soit $x \in \mathbb{R}$ et soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de L^1 . On pose

$$Z_n = x + \sum_{j=1}^n X_j$$

et $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ pour tout $n \in \mathbb{N}_0$. Alors, la suite $(Z_n)_{n \in \mathbb{N}_0}$ est

- une martingale si $\mathbb{E}[X_1] = 0$,
- une sous-martingale si $\mathbb{E}[X_1] \geq 0$,
- une sur-martingale si $\mathbb{E}[X_1] \leq 0$.

Le résultat suivant est très utilisé et permet de construire facilement des sous-martingales à partir de (sous-)martingales.

Proposition 4.5.9

1. Soient $(Z_n)_{n \in \mathbb{N}_0}$ une martingale pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ et f une fonction convexe. Si $f(Z_n) \in L^1$ pour tout $n \in \mathbb{N}_0$, alors $(f(Z_n))_{n \in \mathbb{N}_0}$ est une sous-martingale pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$.
2. Soient $(Z_n)_{n \in \mathbb{N}_0}$ une sous-martingale pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ et f une fonction convexe croissante. Si $f(Z_n) \in L^1$ pour tout $n \in \mathbb{N}_0$, alors $(f(Z_n))_{n \in \mathbb{N}_0}$ est une sous-martingale pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$.

Démonstration : Remarquons que pour tout ensemble borélien $B \subseteq \mathbb{R}$, on a

$$(f(Z_n))^{-1}(B) = \{\omega \in \Omega : Z_n(\omega) \in f^{-1}(B)\} \in \mathcal{F}_n$$

puisque $f^{-1}(B)$ est un ensemble borélien de \mathbb{R} . De plus, dans le premier cas, on applique l'inégalité de Jensen pour les espérances conditionnelles et il vient

$$\mathbb{E}[f(Z_{n+1}) | \mathcal{F}_n] \geq f(\mathbb{E}[Z_{n+1} | \mathcal{F}_n]) = f(Z_n),$$

ce qui suffit. La dernière égalité devient une inégalité dans le deuxième cas. ■

En considérant différents choix de fonctions f , on obtient le corollaire suivant.

Corollaire 4.5.10

1. Si $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale, alors $(|Z_n|)_{n \in \mathbb{N}_0}$ une sous-martingale.
2. Si $p > 1$ et si $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale telle que $\mathbb{E}[|Z_n|^p] < +\infty$, alors $(|Z_n|^p)_{n \in \mathbb{N}_0}$ est une sous-martingale.

Les inégalités de Doob ont de nombreuses applications et permettent notamment d'étudier le supremum d'une suite de variables aléatoires. Nous les présentons ici sous leur forme la plus directe, mais signalons qu'il existe également des inégalités de Doob pour la norme L^p .

Lemme 4.5.11 Soit $(Z_n)_{n \in \mathbb{N}_0}$ une sous-martingale. Alors,

$$\mathbb{E}[Z_m \mathbf{1}_A] \geq \mathbb{E}[Z_n \mathbf{1}_A]$$

pour tout $m \geq n$ et tout $A \in \mathcal{F}_n$.

Démonstration : On sait que $\mathbb{E}[Z_m | \mathcal{F}_n] \geq Z_n$ pour tout $m \geq n$. Par conséquent, pour tout $A \in \mathcal{F}_n$, on a $\mathbb{E}[Z_m \mathbf{1}_A | \mathcal{F}_n] = \mathbf{1}_A \mathbb{E}[Z_m | \mathcal{F}_n] \geq Z_n \mathbf{1}_A$. On en tire alors

$$\mathbb{E}[Z_m \mathbf{1}_A] = \mathbb{E}[\mathbb{E}[Z_m \mathbf{1}_A | \mathcal{F}_n]] \geq \mathbb{E}[Z_n \mathbf{1}_A].$$

■

Proposition 4.5.12 (Inégalités de Doob) Soit $(Z_n)_{n \in \mathbb{N}_0}$ une sous-martingale. Alors, pour tout $a > 0$, on a

$$a \mathbb{P}\left(\max_{k \in \{1, \dots, n\}} Z_k \geq a\right) \leq \mathbb{E}[Z_n \mathbf{1}_{\{\max_{k \in \{1, \dots, n\}} Z_k \geq a\}}]$$

pour tout $n \in \mathbb{N}_0$. Si de plus la suite $(Z_n)_{n \in \mathbb{N}_0}$ est positive, on en tire que

$$\mathbb{P}\left(\max_{k \in \{1, \dots, n\}} Z_k \geq a\right) \leq \frac{\mathbb{E}[Z_n]}{a}$$

pour tout $n \in \mathbb{N}_0$.

Démonstration : Considérons la variable aléatoire T définie par

$$T = \min\{n \in \mathbb{N}_0 : Z_n \geq a\}.$$

Remarquons que

$$\{T \leq k\} = \bigcup_{j=1}^k \underbrace{\{Z_j \geq a\}}_{\in \mathcal{F}_j \subset \mathcal{F}_k} \in \mathcal{F}_k$$

et donc

$$\{T = k\} = \{T \leq k\} \setminus \{T \leq k-1\} \in \mathcal{F}_k.$$

Alors,

$$\mathbb{P}\left(\max_{k \in \{1, \dots, n\}} Z_k \geq a\right) = \mathbb{P}(T \leq n) = \sum_{k=1}^n \mathbb{P}(T = k).$$

De plus, on a

$$a \mathbb{P}(T = k) = a \mathbb{E}[\mathbf{1}_{\{T=k\}}] \leq \mathbb{E}[Z_k \mathbf{1}_{\{T=k\}}] \leq \mathbb{E}[Z_n \mathbf{1}_{\{T=k\}}]$$

par le Lemme 4.5.11. Il s'ensuit que

$$a \mathbb{P}\left(\max_{k \leq n} Z_k \geq c\right) = \sum_{k=1}^n a \mathbb{P}(T = k) \leq \sum_{k=1}^n \mathbb{E}[Z_n \mathbf{1}_{\{T=k\}}] = \mathbb{E}\left[Z_n \sum_{k=1}^n \mathbf{1}_{\{T=k\}}\right] = \mathbb{E}[Z_n \mathbf{1}_{\{T \leq n\}}].$$

La deuxième inégalité est évidente. \blacksquare

L'intérêt de cette inégalité est qu'elle permet de majorer une quantité faisant intervenir tout le processus jusqu'au temps n par une quantité ne dépendant que de Z_n .

Remarque 4.5.13 En particulier, si $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale, alors pour tout $a > 0$, on a

$$\mathbb{P}\left(\sup_{k \leq n} |Z_k| \geq a\right) \leq \frac{\mathbb{E}[|Z_n|]}{a}.$$

Les inégalités de Doob permettent de retrouver l'inégalité de Kolmogorov présentée dans le Théorème 5.4.5.

Proposition 4.5.14 (Inégalité de Kolmogorov) Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires indépendantes et centrées de L^2 . Pour tout $a > 0$, on a

$$\mathbb{P}\left(\sup_{n \in \mathbb{N}_0} \left|\sum_{k=1}^n X_k\right| > a\right) \leq \frac{1}{a^2} \sum_{k=1}^{+\infty} \mathbb{E}[X_k^2].$$

Démonstration : Pour tout $n \in \mathbb{N}_0$, on pose

$$Z_n = \sum_{k=1}^n X_k.$$

On sait que la suite $(Z_n)_{n \in \mathbb{N}_0}$ définit une martingale. Par conséquent, $(Z_n^2)_{n \in \mathbb{N}_0}$ est une sous-martingale. En utilisant l'inégalité de Doob, on obtient

$$\mathbb{P}\left(\max_{j \in \{1, \dots, n\}} |Z_j| > a\right) \leq \mathbb{P}\left(\max_{j \in \{1, \dots, n\}} |Z_j| \geq a\right) = \mathbb{P}\left(\max_{j \in \{1, \dots, n\}} Z_j^2 \geq a^2\right) \leq \frac{1}{a^2} \mathbb{E}[Z_n^2] = \frac{1}{a^2} \sum_{k=1}^n \mathbb{E}[X_k^2]$$

pour tout $n \in \mathbb{N}_0$. On en tire que

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\max_{j \in \{1, \dots, n\}} |Z_j| > a\right) \leq \frac{1}{a^2} \sum_{k=1}^{+\infty} \mathbb{E}[X_k^2].$$

Remarquons à présent que la suite d'événements $(\{\max_{j \in \{1, \dots, n\}} |Z_j| > a\})_{n \in \mathbb{N}}$ est croissante et donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\max_{j \in \{1, \dots, n\}} |Z_j| > a\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}_0} \left\{\max_{j \in \{1, \dots, n\}} |Z_j| > a\right\}\right) = \mathbb{P}\left(\sup_{n \in \mathbb{N}_0} |Z_n| > a\right),$$

ce qui permet de conclure. \blacksquare

4.6 Temps d'arrêt et martingales arrêtées

Une martingale $(Z_n)_{n \in \mathbb{N}_0}$ sert principalement à décrire l'évolution d'un processus au cours du temps (Z_n est alors l'état du système à l'instant n). A tout processus évolutif, on associe la définition suivante.

Définition 4.6.1 Soit $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ une filtration. Une variable aléatoire T est un *temps d'arrêt* pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ si

1. T est à valeurs dans $\mathbb{N} \cup \{+\infty\}$,
2. $\{T \leq n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}_0$.

Un temps d'arrêt est donc un temps aléatoire qui suit l'évolution aléatoire sous-jacente. Dans notre exemple d'un jeu aléatoire répété, T peut être vu comme le temps auquel le joueur décide de quitter le jeu : cette décision de quitter le jeu directement après le temps n ne dépend que des résultats obtenus avant le temps n (inclus).

Remarque 4.6.2 La condition $\{T \leq n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}_0$ est équivalente à $\{T = n\} \in \mathcal{F}_n$. En effet, d'une part, on a

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n - 1\}$$

et d'autre part

$$\{T \leq n\} = \bigcup_{k=0}^n \{T = k\}.$$

Notons que cette définition ne se généralise pas au cas continu que nous aborderons par la suite.

Exemple 4.6.3 Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires réelles adaptées à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$. Pour tout $a > 0$, le temps aléatoire

$$T_a = \inf\{n \in \mathbb{N}_0 : X_n > a\},$$

appelé *temps de premier passage au seuil a* , est un temps d'arrêt. En effet, on a

$$\{T_a \leq n\} = \{\exists m \leq n : X_m > a\} = \bigcup_{m=1}^n \{X_m > a\} \in \mathcal{F}_n.$$

D'une manière générale, le temps de première entrée dans un borélien B , défini par

$$T_B = \inf\{n \in \mathbb{N}_0 : X_n \in B\},$$

est un temps d'arrêt. Par contre, le temps de dernier passage au seuil a , défini par

$$\tilde{T}_a = \sup\{n \in \mathbb{N}_0 : X_n > a\},$$

n'est pas un temps d'arrêt.

Le théorème suivant sera la base de cette section. On note $T \wedge n = \min\{T, n\}$.

Théorème 4.6.4 Soit $(Z_n)_{n \in \mathbb{N}_0}$ une martingale (resp. sous-martingale, sur-martingale) pour une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$. Si T un temps d'arrêt pour $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$, alors

$$Z_{T \wedge n}, \quad n \in \mathbb{N}_0$$

est une martingale (resp. sous-martingale, sur-martingale) pour $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$. En particulier, pour tout $n \in \mathbb{N}_0$, on a

$$\mathbb{E}[Z_{T \wedge n}] = \mathbb{E}[Z_1]$$

(resp. $\mathbb{E}[Z_{T \wedge n}] \geq \mathbb{E}[Z_1]$, $\mathbb{E}[Z_{T \wedge n}] \leq \mathbb{E}[Z_1]$).

Reprenons notre exemple du jeu de hasard répété. On peut décider d'arrêter le jeu lorsqu'on a gagné un certain montant A ou lorsqu'on aura joué 10 parties ($n = 10$). On considère donc le temps d'arrêt $T = \inf\{n \in \mathbb{N}_0 : Z_n = A\}$ (on montrera par la suite qu'il est fini presque sûrement) et le gain espéré est alors

$$\mathbb{E}[Z_{T \wedge 10}] = \mathbb{E}[Z_1].$$

Par contre, si on joue jusqu'à avoir gagné un certain montant A , l'espérance de gain est donnée par

$$\mathbb{E}[Z_T] = A.$$

Par conséquent, même si $\mathbb{E}[Z_{T \wedge n}] = \mathbb{E}[Z_1]$, il est possible de construire une martingale $(Z_n)_{n \in \mathbb{N}_0}$ et un temps d'arrêt T tels que

$$\mathbb{E}[Z_T] \neq \mathbb{E}[Z_1].$$

Si T est bornée, alors pour tout $n \geq M$, on a $Z_{T \wedge n} = Z_T$. Il s'ensuit que $\mathbb{E}[Z_T] = \mathbb{E}[Z_{T \wedge n}] = \mathbb{E}[Z_1]$ par le Théorème 4.6.4.

Si T n'est pas bornée, pour que Z_T soit bien défini on doit s'assurer que $\mathbb{P}[T < +\infty] = 1$. Dans ce cas, on a presque sûrement $Z_{n \wedge T} = Z_T$ pour tout n suffisamment grand et donc

$$\lim_{n \rightarrow \infty} Z_{n \wedge T} = Z_T,$$

d'où

$$\mathbb{E}\left[\lim_{n \rightarrow \infty} Z_{n \wedge T}\right] = \mathbb{E}[Z_T].$$

De plus, par le Théorème 4.6.4, on sait que

$$\mathbb{E}[Z_{n \wedge T}] = \mathbb{E}[Z_1]$$

et donc si on peut intervertir la limite et l'espérance, on aura $\mathbb{E}[Z_T] = \mathbb{E}[Z_1]$. C'est le cas par exemple par le théorème de la convergence dominée si il existe une constante déterministe $M > 0$ telle que $|Z_{n \wedge T}| \leq M$ pour tout $n \in \mathbb{N}_0$.

Terminons cette section par une application au problème de la ruine pour deux joueurs. Soient $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d Rademacher, i.e.

$$\mathbb{P}(X_n = 1) = 1/2 = \mathbb{P}(X_n = -1).$$

Posons $M_0 = 0$, $M_n = X_1 + \dots + X_n$ et définissons

$$T = \min\{n : M_n = A \text{ ou } M_n = -B\}.$$

pour $A, B \geq 0$ entiers.

Proposition 4.6.5 *Presque sûrement, $T < \infty$ et on a $\mathbb{P}(M_T = A) = \frac{B}{A+B}$.*

Démonstration : Remarquons que, quelle que soit la position de la promenade aléatoire, une suite de $A + B$ “+1” suffira pour garantir que la barrière $y = A$ sera dépassée. Donc

$$\mathbb{P}(T > n(A + B)) \leq \mathbb{P}(E_1^c \cap \dots \cap E_n^c)$$

avec

$$E_k = \{X_{k(A+B)} = X_{k(A+B)+1} = \dots = X_{(k+1)(A+B)-1}\}.$$

Les événements E_k étant indépendants (ils concernent des indices qui ne se croisent pas) on calcule

$$\mathbb{P}(T > n(A + B)) \leq \left(1 - \left(\frac{1}{2}\right)^{A+B}\right)^n.$$

Par conséquent

$$\mathbb{P}(T = \infty) \leq \lim_{n \rightarrow \infty} \mathbb{P}(T > n(A + B)) \leq \lim_{n \rightarrow \infty} \left(1 - \left(\frac{1}{2}\right)^{A+B}\right)^n = 0.$$

La première affirmation est donc prouvée. Vu que $T < \infty$ presque sûrement, nous pouvons déduire

$$\mathbb{E}[M_T] = \mathbb{E}[M_0] = 0$$

puisque $|M_{T \wedge n}| \leq \max\{A, B\}$. Puisque

$$\begin{aligned} \mathbb{E}[M_T] &= AP(M_T = A) - BP(M_T = -B) \\ &= (A + B)P(M_T = A) - B, \end{aligned}$$

la conclusion s’ensuit. ■

Remarque 4.6.6 Ce problème est souvent connu sous le nom de Gambler’s ruin, appellation qui peut se justifier comme suit. Considérons un joueur atavique, compulsif et excessivement riche, jouant plusieurs parties d’un jeu équilibré. S’il mise un euro à chaque partie, le processus S_n représentera sa fortune après les n parties. La question ci-dessus devient alors quelle est la probabilité que notre sympathique joueur gagne A euros avant d’en perdre B ?

Pour montrer qu’un temps d’arrêt T est presque sûrement fini, il est commun d’utiliser les propriétés de convergence des martingales. Les martingales ont l’élégance de ne pas exiger grand-chose pour converger. Remarquons que la définition d’une sous-martingale est semblable à celle d’une suite qui, en tendance et conditionnellement au passé, est croissante. Nous savons qu’une suite de réels qui est croissante et majorée converge. Nous allons montrer un résultat analogue pour les sous-martingales, en remplaçant l’hypothèse de suite majorée par une majoration de l’espérance, i.e. en imposant $\sup_n \mathbb{E}[|Z_n|] < +\infty$.

Théorème 4.6.7 (Convergence presque sûre) Soit $(Z_n)_{n \in \mathbb{N}_0}$ une sous-martingale. Supposons que

$$\sup_{n \in \mathbb{N}_0} \mathbb{E}[Z_n^+] < +\infty.$$

Alors $(Z_n)_{n \in \mathbb{N}_0}$ converge presque sûrement vers une variable aléatoire $Z \in L^1$.

Corollaire 4.6.8 Si $(Z_n)_{n \in \mathbb{N}_0}$ est une martingale positive, alors $(Z_n)_{n \in \mathbb{N}_0}$ converge presque sûrement vers une variable aléatoire finie.

Exemple 4.6.9 Considérons le problème de la ruine du joueur participant à un jeu équitable : Z_n représente le gain obtenu après le $n^{\text{ième}}$ jeu. Supposons en outre que la banque ne fait pas crédit, donc à chaque partie, le joueur gagne ou perd une pièce, mais le jeu s'arrête si le joueur perd alors qu'il ne lui restait pas d'argent. Soit T le temps d'arrêt égal au nombre de parties jouées avant que le joueur ne perde tout son argent, i.e. $T = \{n \geq 1 : Z_n = 0\}$. Par le Théorème 4.6.4, $(Z_{n \wedge T})_{n \in \mathbb{N}_0}$ est une martingale et puisqu'elle est positive, la limite

$$\lim_{n \rightarrow +\infty} Z_{n \wedge T}$$

existe et est finie presque sûrement par le Corollaire 4.6.8. Or, il est clair que pour tout $n \in \mathbb{N}_0$, on a $|Z_{n+1} - Z_n| = 1$. Par conséquent, si $T = +\infty$ avec une probabilité non-nulle, alors la suite $(Z_{n \wedge T})_{n \in \mathbb{N}_0}$ ne converge pas vers une variable finie sur un événement de probabilité non-nulle, ce qui est impossible. On en tire donc que $\mathbb{P}(T < +\infty) = 1$ et le joueur finira par perdre tout son argent.

Chapitre 5

Convergences

Dans ce chapitre, nous considérons une suite de variables aléatoires ou de vecteurs aléatoires $(X_n)_{n \in \mathbb{N}}$ et nous nous intéressons au comportement asymptotique de la suite.

5.1 Convergence presque sûre

Le premier mode de convergence que nous étudions est la convergence presque sûre. Nous supposons que les vecteurs aléatoires avec lesquels nous travaillons sont tous définis sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Pour simplifier les notations, on suppose ici travailler avec des vecteurs aléatoires de \mathbb{R}^d écrits en ligne $X = (X_1, \dots, X_d)$.

Définition 5.1.1 Une suite $(X_n)_{n \in \mathbb{N}}$ de vecteurs aléatoires de \mathbb{R}^d définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ converge \mathbb{P} -presque sûrement vers le vecteur aléatoire X si

$$\mathbb{P} \left(\lim_{n \rightarrow +\infty} X_n = X \right) = 1.$$

Dans ce cas, on écrit $X_n \xrightarrow{\text{p.s.}} X$ ou $\lim_{n \rightarrow +\infty} X_n = X$ p.s.

On parle aussi de convergence “forte” par opposition à d’autres modes de convergence dits “faibles”, ou de convergence “presque partout” ou encore “avec probabilité 1”. La convergence presque sûre revient à demander que l’ensemble des $\omega \in \Omega$ tels que $X_n(\omega)$ converge vers $X(\omega)$ (au sens classique d’une suite de vecteurs de \mathbb{R}^d) est de mesure 1 par rapport à \mathbb{P} . Remarquons qu’on sait que cet ensemble est bien un événement de \mathcal{F} puisque

$$\left\{ \omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \right\} = \bigcap_{p \in \mathbb{N}_0} \bigcup_{m \in \mathbb{N}_0} \bigcap_{n \geq m} \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| < \frac{1}{p} \right\}$$

par définition de la convergence d’une suite. Les résultats connus sur la convergence ponctuelle permettent directement d’obtenir les propriétés suivantes.

Lemme 5.1.2 Soient X, Y des vecteurs aléatoires de \mathbb{R}^d et soient $(X_n)_{n \in \mathbb{N}}, (Y_n)_{n \in \mathbb{N}}$ des suites de vecteurs aléatoires de \mathbb{R}^d . Si $X_n = (X_{n,1}, \dots, X_{n,d})$ et $X = (X_1, \dots, X_d)$, alors $X_n \xrightarrow{\text{p.s.}} X$ si et seulement si $X_{n,j} \xrightarrow{\text{p.s.}} X_j$ pour tout $j \in \{1, \dots, d\}$. De plus, si $X_n \xrightarrow{\text{p.s.}} X$ et $Y_n \xrightarrow{\text{p.s.}} Y$, alors

- si $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$ est une application continue, alors $f(X_n) \xrightarrow{\text{p.s.}} f(X)$,

- pour tous $\alpha, \beta \in \mathbb{R}$, on a $\alpha X_n + \beta Y_n \xrightarrow{p.s.} \alpha X + \beta Y$,
- $X_n Y_n^T \xrightarrow{p.s.} X Y^T$,
- $(X_n, Y_n) \xrightarrow{p.s.} (X, Y)$.

Le lemme suivant donne des critères équivalents pour la convergence presque sûre.

Lemme 5.1.3 Soient $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires de \mathbb{R}^d définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et X un vecteur aléatoire de \mathbb{R}^d défini sur $(\Omega, \mathcal{F}, \mathbb{P})$. Les affirmations suivantes sont équivalentes :

1. $X_n \xrightarrow{p.s.} X$,
2. pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| < \varepsilon \quad \forall n \geq m) \rightarrow 1$ lorsque $m \rightarrow +\infty$,
3. pour tout $\varepsilon > 0$, $\mathbb{P}(\sup_{n \geq m} |X_n - X| \geq \varepsilon) \rightarrow 0$ lorsque $m \rightarrow +\infty$.

Démonstration : 1 \Leftrightarrow 2. Pour tous $\varepsilon > 0$ et $m \in \mathbb{N}_0$, on pose

$$A_{m,\varepsilon} = \bigcap_{n \geq m} \{|X_n - X| < \varepsilon\}.$$

La continuité à droite de la mesure de probabilité \mathbb{P} donne

$$\mathbb{P} \left(\bigcap_{p \in \mathbb{N}_0} \bigcup_{m \in \mathbb{N}_0} A_{m,1/p} \right) = \lim_{p \rightarrow +\infty} \mathbb{P} \left(\bigcup_{m \in \mathbb{N}_0} A_{m,1/p} \right).$$

Par conséquent, on a $X_n \xrightarrow{p.s.} X$ si et seulement si

$$\lim_{p \rightarrow +\infty} \mathbb{P} \left(\bigcup_{m \in \mathbb{N}_0} A_{m,1/p} \right) = 1.$$

Comme la suite $(\bigcup_{m \in \mathbb{N}_0} A_{m,1/p})_p$ est décroissante en p , la suite $(\mathbb{P}(\bigcup_{m \in \mathbb{N}_0} A_{m,1/p}))_p$ de $[0, 1]$ est décroissante en p . Ainsi, sa limite vaut 1 si et seulement si

$$\mathbb{P} \left(\bigcup_{m \in \mathbb{N}_0} A_{m,1/p} \right) = 1 \quad \forall p \in \mathbb{N}_0.$$

Puisque pour tout $0 < \varepsilon < 1$, il existe $p \in \mathbb{N}_0$ tel que $\frac{1}{p+1} < \varepsilon \leq \frac{1}{p}$, il suit que $X_n \xrightarrow{p.s.} X$ si et seulement si

$$\mathbb{P} \left(\bigcup_{m \in \mathbb{N}_0} A_{m,\varepsilon} \right) = 1 \quad \forall \varepsilon > 0.$$

Clairement, pour chaque $\varepsilon > 0$, la suite $(A_{m,\varepsilon})_m$ est croissante en m , donc la continuité à droite de la mesure \mathbb{P} permet d'écrire

$$\mathbb{P} \left(\bigcup_{m \in \mathbb{N}_0} A_{m,\varepsilon} \right) = \lim_{m \rightarrow +\infty} \mathbb{P}(A_{m,\varepsilon}),$$

ce qui nous permet de conclure.

2 ⇔ 3. Reprenons les mêmes notations que ci-dessus. Alors, par passage au complémentaire, on a que $\lim_{m \rightarrow +\infty} \mathbb{P}(A_{m,1/\varepsilon}) = 1$ pour tout $\varepsilon > 0$ si et seulement si $\lim_{m \rightarrow +\infty} \mathbb{P}(A_{m,1/\varepsilon}^c) = 0$ pour tout $\varepsilon > 0$. Remarquons à présent que

$$A_{m,1/\varepsilon}^c = \bigcup_{n \geq m} \{|X - X_n| \geq \varepsilon\} = \left\{ \sup_{n \geq m} |X - X_n| \geq \varepsilon \right\},$$

d'où la conclusion. ■

Remarque 5.1.4 Les convergences des suites de probabilités des items 2 et 3 du Lemme 5.1.3 étant valables pour tout $\varepsilon > 0$, on peut évidemment de manière équivalente remplacer les inégalités strictes par des inégalités non-strictes.

Bien sûr, pour montrer que la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement, il suffit de montrer qu'elle est presque sûrement de Cauchy, c'est-à-dire que l'événement

$$\{\omega \in \Omega : (X_n(\omega))_{n \in \mathbb{N}} \text{ est de Cauchy}\} = \bigcap_{p \in \mathbb{N}_0} \bigcup_{m \in \mathbb{N}_0} \bigcap_{n, k \geq m} \left\{ \omega \in \Omega : |X_n(\omega) - X_k(\omega)| < \frac{1}{p} \right\}$$

est de probabilité 1.

Le Lemme de Borel-Cantelli permet de démontrer un critère de convergence presque sûre pratique.

Proposition 5.1.5 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires de \mathbb{R}^d définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et X un vecteur aléatoire de \mathbb{R}^d défini sur $(\Omega, \mathcal{F}, \mathbb{P})$.

- Si on a

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$$

pour tout $\varepsilon > 0$, alors $X_n \xrightarrow{p.s.} X$.

- Si les vecteurs aléatoires $X_n, n \in \mathbb{N}$, sont indépendants, alors $X_n \xrightarrow{p.s.} 0$ si et seulement si

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) < +\infty$$

pour tout $\varepsilon > 0$.

Démonstration : Pour le premier point, le lemme de Borel-Cantelli et la convergence de la série impliquent que

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} \{|X_n - X| > \varepsilon\} \right) = \mathbb{P} \left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - X| > \varepsilon\} \right) = 0.$$

La continuité à droite de \mathbb{P} implique que

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\bigcup_{n \geq m} \{|X_n - X| > \varepsilon\} \right) = 0,$$

ce qui suffit par la Proposition 5.1.3 puisque $\bigcup_{n \geq m} \{|X_n - X| > \varepsilon\} = \{\sup_{n \geq m} |X_n - X| > \varepsilon\}$.

Pour le second point, on remarque que si les $X_n, n \in \mathbb{N}$, sont indépendants, alors les événements $\{|X_n| > \varepsilon\}, n \in \mathbb{N}$, sont indépendants et on utilise à nouveau le lemme de Borel-Cantelli pour conclure. ■

Remarque 5.1.6 On est obligé de supposer $X = 0$ dans le deuxième point, sans quoi on n'a pas l'indépendance des événements.

5.2 Convergence en moyenne et en norme L^p

Rappelons que les espaces L^p peuvent être munis de la norme $\|\cdot\|_p$ donnée par

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$$

pour tout $X \in L^p$. On peut donc, en particulier, définir une notion de convergence. Rappelons que les espaces L^p munis de leur norme sont des espaces complets (de Banach). Comme un vecteur aléatoire est dans L^p si chacune de ses composantes est dans L^p , on va travailler uniquement avec des variables aléatoires.

Définition 5.2.1 Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires de L^p définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ converge dans L^p ou en norme L^p vers la variable aléatoire X si

$$\lim_{n \rightarrow +\infty} \|X_n - X\|_p = 0.$$

Dans ce cas, on écrit $X_n \xrightarrow{p} X$. Lorsque $p = 1$, on parle de *convergence en moyenne*. Lorsque $p = 2$, on parle de *convergence en moyenne quadratique* et on écrit $X_n \xrightarrow{qm} X$.

Remarquons que si elle existe, la limite en norme L^p est presque sûrement unique puisque si $X_n \xrightarrow{p} X$ et $X_n \xrightarrow{p} Y$, alors $\|X - Y\|_p = 0$. Le Corollaire 2.5.7 implique que si $p \geq q \geq 1$ et si $X_n \xrightarrow{p} X$, alors $X_n \xrightarrow{q} X$. En particulier, la convergence en moyenne quadratique implique la convergence en moyenne.

En général, il n'y a pas de lien entre la notion de convergence en norme L^p et la convergence presque sûre.

Exemple 5.2.2 Considérons l'espace $[0, 1]$ muni de la mesure de Lebesgue et la suite $(X_n)_{n \in \mathbb{N}_0}$ définie par $X_n = n \mathbb{1}_{[0, 1/n]}$. Clairement, cette suite converge presque partout vers 0. Néanmoins, $(X_n)_{n \in \mathbb{N}_0}$ ne converge pas vers 0 en moyenne puisque

$$\mathbb{E}[|X_n|] = n\lambda([0, 1/n]) = 1$$

pour tout $n \in \mathbb{N}_0$. Comme la convergence en norme L^p pour $p > 1$ implique la convergence en moyenne, on n'a pas non plus la convergence en norme L^p .

Exemple 5.2.3 Considérons l'espace $[0, 1]$ muni de la mesure de Lebesgue et la suite $(X_n)_{n \in \mathbb{N}_0}$ définie par $X_1 = \mathbb{1}_{[0, 1]}$, $X_2 = \mathbb{1}_{[0, 1/2]}$, $X_3 = \mathbb{1}_{[1/2, 1]}$, et de manière générale, si $n = 2^j + k$ avec $k \in \{0, \dots, 2^j - 1\}$,

$$X_n = \mathbb{1}_{[k/2^j, (k+1)/2^j]}$$

Alors, la suite $(X_n)_{n \in \mathbb{N}_0}$ converge en norme L^p vers 0 puisqu'on a pour $n = 2^j + k$,

$$\mathbb{E}[|X_n|^p] = \mathbb{E}[|X_n|] = \lambda([k/2^j, (k+1)/2^j]) = 2^{-j}.$$

Par contre, la suite $(X_n)_{n \in \mathbb{N}_0}$ ne converge pas presque sûrement vers 0 car pour tout $x \in [0, 1[$, la suite $(X_n(x))_{n \in \mathbb{N}_0}$ prend une infinité de fois la valeur 0 et une infinité de fois la valeur 1.

Cependant, il existe des hypothèses supplémentaires sous lesquelles la convergence presque sûre implique la convergence en norme L^p .

Proposition 5.2.4 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables telle que $X_n \xrightarrow{p.s.} X$. S'il existe une variable aléatoire positive $Z \in L^1$ telle que $|X_n|^p \leq Z$ presque sûrement pour tout $n \in \mathbb{N}$, alors $X_n \xrightarrow{p} X$.*

Démonstration : Bien sûr, on a également $|X|^p \leq Z$ presque sûrement, d'où

$$|X_n - X|^p \leq (|X_n| + |X|)^p \leq 2^p |Z|$$

presque sûrement. Le résultat est alors une conséquence du théorème de la convergence dominée. ■

Le résultat suivant sera obtenu comme conséquence directe de résultats de la section suivante. Nous en donnons ici une preuve alternative.

Proposition 5.2.5 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables de L^p . Si $X_n \xrightarrow{p} X$, alors il existe une sous-suite $(n_k)_{k \in \mathbb{N}}$ telle que $X_{n_k} \xrightarrow{p.s.} X$.*

Démonstration : Comme la convergence en norme L^p implique la convergence en norme L^1 , on a $X_n \xrightarrow{1} X$. Alors, il existe n_1 tel que

$$\|X_{n_1} - X\|_1 \leq \frac{1}{2}.$$

Par induction, si l'entier n_{k-1} a été construit, on fixe $n_k > n_{k-1}$ tel que

$$\|X_{n_k} - X\|_1 \leq \frac{1}{2^k}.$$

Montrons que la suite $(X_{n_k})_{k \in \mathbb{N}_0}$ ainsi construite converge presque sûrement vers X . Si pour tout $k \in \mathbb{N}_0$ on pose

$$Y_k = |X_{n_k} - X|,$$

alors $\mathbb{E}[Y_k] \leq 2^{-k}$ et donc par le théorème de convergence monotone,

$$\mathbb{E}\left[\sum_{k \in \mathbb{N}_0} Y_k\right] \leq \sum_{k \in \mathbb{N}_0} 2^{-k} < +\infty.$$

Il s'ensuit que $\sum_{k \in \mathbb{N}_0} Y_k$ est presque sûrement fini. En particulier, on trouve que $Y_k \xrightarrow{p.s.} 0$, c'est-à-dire $X_{n_k} \xrightarrow{p.s.} X$. ■

Nous en déduisons une conséquence importante d'unicité de la limite.

Corollaire 5.2.6 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables de L^p . Si $X_n \xrightarrow{p.s.} X$ et $X_n \xrightarrow{p} Y$, alors $X = Y$ presque sûrement.

5.3 Convergence en probabilité

Introduisons à présent un troisième mode de convergence pour les variables aléatoires.

Définition 5.3.1 Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ converge *en probabilité* vers la variable aléatoire X si

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

pour tout $\varepsilon > 0$. Dans ce cas, on écrit $X_n \xrightarrow{\mathbb{P}} X$.

Exemple 5.3.2 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes (ou simplement non corrélées) centrées et de variance $\text{Var}[X_n] = \sigma^2$ pour tout $n \in \mathbb{N}$. Alors, pour tout $\varepsilon > 0$, l'inégalité de Tchebychev donne

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{j=1}^n X_j \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

qui converge vers 0 lorsque n tend vers l'infini. En particulier, on déduit que

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\mathbb{P}} 0.$$

Remarque 5.3.3 La convergence en probabilité peut être définie par les deux distances suivantes sur l'espace des variables aléatoires :

$$d_1(X, Y) = \mathbb{E} \left(\frac{|X - Y|}{1 + |X - Y|} \right)$$

$$d_2(X, Y) = \inf \{ \varepsilon > 0 : \mathbb{P}(|X - Y| \geq \varepsilon) \leq \varepsilon \}.$$

Pour ces distances, l'espace des variables aléatoires est un espace vectoriel topologique (non localement convexe) complet.

Le résultat suivant montre qu'on a unicité de la limite en probabilité.

Proposition 5.3.4 Si $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\mathbb{P}} Y$, alors $X = Y$ presque sûrement.

Démonstration : Puisque $|X - Y| \leq |X - X_n| + |Y - X_n|$, on a

$$\{|X - Y| \geq \varepsilon\} \subseteq \{|X - X_n| \geq \varepsilon/2\} \cup \{|Y - X_n| \geq \varepsilon/2\}$$

pour tout $\varepsilon > 0$. On en tire que

$$\mathbb{P}(|X - Y| \geq \varepsilon) \leq \mathbb{P}(|X - X_n| \geq \varepsilon/2) + \mathbb{P}(|Y - X_n| \geq \varepsilon/2)$$

et en faisant tendre n vers l'infini,

$$\mathbb{P}(|X - Y| \geq \varepsilon) = 0.$$

Pour conclure, il suffit alors de remarquer que

$$\{X \neq Y\} = \bigcup_{k \in \mathbb{N}_0} \{|X - Y| \geq 1/k\}.$$

■

Intéressons-nous à présent aux relations qui existent entre la convergence en probabilité et les deux types de convergence déjà considérés. On remarque aisément la différence avec la convergence presque sûre qui a des exigences quant au comportement du supremum vu le Lemme 5.1.3. On peut le réécrire de la manière suivante.

Lemme 5.3.5 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Alors, $X_n \xrightarrow{p.s.} X$ si et seulement si*

$$Y_n := \sup_{k \geq n} |X_k - X| \xrightarrow{\mathbb{P}} 0.$$

Ainsi, la convergence presque sûre peut être vue comme une convergence uniforme en probabilité.

Théorème 5.3.6 *Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et X une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$.*

- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.
- On a $X_n \xrightarrow{\mathbb{P}} X$ si et seulement si de toute suite déterministe croissante d'entiers (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k} \xrightarrow{p.s.} X$.

Démonstration : Le premier point est une conséquence immédiate du Lemme 5.3.5.

Pour le second point, supposons que $X_n \xrightarrow{\mathbb{P}} X$. Si (n') est une suite croissante d'entiers, on a évidemment également $X_{n'} \xrightarrow{\mathbb{P}} X$. Ainsi, il existe n'_1 tel que

$$\mathbb{P}\left(|X_{n'_1} - X| > 1\right) \leq \frac{1}{2}.$$

Par induction, si l'entier n'_{k-1} a été défini, on peut trouver un indice $n'_k > n'_{k-1}$ tel que

$$\mathbb{P}\left(|X_{n'_k} - X| > \frac{1}{k}\right) \leq \frac{1}{2^k}.$$

Montrons que la suite $(X_{n'_k})_{k \in \mathbb{N}_0}$ ainsi construite converge presque sûrement vers X en utilisant la Proposition 5.1.5. Soit $\varepsilon > 0$ et considérons $j \in \mathbb{N}_0$ tel que $\frac{1}{j} < \varepsilon$. Alors, on a

$$\sum_{k=j}^{+\infty} \mathbb{P}\left(|X_{n'_k} - X| > \varepsilon\right) \leq \sum_{k=j}^{+\infty} \mathbb{P}\left(|X_{n'_k} - X| > \frac{1}{k}\right) \leq \sum_{k=j}^{+\infty} \frac{1}{2^k} < +\infty,$$

ce qui suffit. Pour la réciproque, supposons que $(X_n)_{n \in \mathbb{N}}$ ne converge pas en probabilité vers X . Alors, il existe $\varepsilon > 0$ et $\eta > 0$ tels que

$$\mathbb{P}(|X_{n'} - X| \geq \varepsilon) \geq \eta$$

pour une sous-suite (n') . Par hypothèse, il existe une sous-suite (n'_k) telle que $X_{n'_k} \xrightarrow{\text{p.s.}} X$. Le premier point implique alors que la convergence a lieu également en probabilité, et donc

$$\lim_{k \rightarrow +\infty} \mathbb{P}(|X_{n'_k} - X| \geq \varepsilon) = 0,$$

ce qui contredit le choix de la sous-suite (n') . ■

Corollaire 5.3.7 Si $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\text{p.s.}} Y$, alors $X = Y$ presque sûrement.

On sait que la convergence presque sûre implique la convergence en probabilité, mais l'inverse n'est pas vrai comme le montrent les exemples suivants.

Exemple 5.3.8 Reprenons les variables aléatoires de l'exemple 5.2.3. On sait que la suite $(X_n)_{n \in \mathbb{N}_0}$ ne converge pas presque sûrement. Par contre, $X_n \xrightarrow{\mathbb{P}} X$ puisque si $n = 2^j + k$, on a

$$\mathbb{P}(|X_n| > \varepsilon) = \lambda([k/2^j, (k+1)/2^j[) = 2^{-j}$$

qui tend vers 0 lorsque n (et donc j) tend vers l'infini.

Exemple 5.3.9 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes de loi Bern(p_n). Alors, $X_n \xrightarrow{\mathbb{P}} 0$ si et seulement si $p_n \rightarrow 0$. Mais la Proposition 5.1.5 implique que $X_n \xrightarrow{\text{p.s.}} 0$ si et seulement si $\sum_{n \in \mathbb{N}} p_n < +\infty$, ce qui est une condition strictement plus forte.

Grâce à ce résultat on obtient quasiment gratuitement le résultat suivant.

Lemme 5.3.10 On peut remplacer toutes les flèches $\xrightarrow{\text{p.s.}}$ par des flèches $\xrightarrow{\mathbb{P}}$ dans le Lemme 5.1.2.

Intéressons-nous à présent aux liens qui existent entre la convergence en probabilité et la convergence en norme L^p .

Proposition 5.3.11

- Si $X_n \xrightarrow{p} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.
- S'il existe une variable aléatoire positive $Z \in L^1$ telle que $|X_n|^p \leq Z$ presque sûrement pour tout $n \in \mathbb{N}$ et si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{p} X$ pour tout $p \geq 1$.

Démonstration : Tout d'abord, l'inégalité de Markov donne

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p}$$

pour tout $\varepsilon > 0$, qui tend vers 0 lorsque n tend vers l'infini. Cela démontre le premier point.

Pour le deuxième, supposons que $(X_n)_{n \in \mathbb{N}}$ ne converge pas en norme L^p vers X . Alors, il existe $\varepsilon > 0$ et une sous-suite (n') de $(n_k)_{k \in \mathbb{N}}$ telle que

$$\|X_{n'_k} - X\|_p \geq \varepsilon.$$

Le Théorème 5.3.6 permet d'extraire une sous-suite $(n'_k)_{k \in \mathbb{N}}$ telle que $X_{n'_k} \xrightarrow{\text{p.s.}} X$. La Proposition 5.2.4 implique alors que $X_{n'_k} \xrightarrow{p} X$, ce qui contredit le choix de la suite (n') . ■

Remarque 5.3.12 Puisque la convergence en norme L^p implique la convergence en probabilité, le Théorème 5.3.6 donne l'existence d'une sous-suite pour laquelle on a la convergence presque sûre. En particulier, cela prouve la Proposition 5.2.5.

A nouveau, la réciproque est fautive en général : la convergence en probabilité n'implique pas la convergence en norme L^p .

Exemple 5.3.13 Reprenons les variables aléatoires de l'exemple 5.2.2. On sait que la suite $(X_n)_{n \in \mathbb{N}_0}$ ne converge pas en norme L^p quel que soit $p \geq 1$. Néanmoins, la suite converge vers 0 en probabilité puisque pour tout $\varepsilon \in]0, 1[$, on a

$$\mathbb{P}(|X_n| > \varepsilon) = \lambda([0, 1/n]) = \frac{1}{n}$$

qui tend vers 0 lorsque n tend vers l'infini.

5.4 Lois des grands nombres

Dans cette section, nous nous intéressons au comportement asymptotique de la somme

$$S_n := \sum_{j=1}^n X_j$$

lorsque $(X_n)_{n \in \mathbb{N}_0}$ est une suite de variables aléatoires. Remarquons que S_n/n n'est rien d'autre que la moyenne empirique de l'échantillon (X_1, \dots, X_n) . Si les X_n sont i.i.d. de loi commune $X \in L^2$, on pose

$$\bar{X}^{(n)} := \frac{1}{n} \sum_{j=1}^n X_j.$$

On calcule directement que

$$\mathbb{E}[\bar{X}^{(n)}] = \frac{1}{n} \sum_{j=1}^n \mathbb{E}[X_j] = \mathbb{E}[X]$$

et l'indépendance des variables aléatoires donne

$$\text{Var}[\bar{X}^{(n)}] = \frac{1}{n^2} \sum_{j=1}^n \text{Var}[X_j] = \frac{\text{Var}[X]}{n}.$$

Ainsi, la variable aléatoire $\bar{X}^{(n)}$ est centrée en $\mathbb{E}[X]$ et plus n augmente, plus sa variance diminue.

On a en fait $\bar{X}^{(n)} \xrightarrow{\mathbb{P}} \mathbb{E}[X]$ car

$$\mathbb{P}\left(\left|\bar{X}^{(n)} - \mathbb{E}[X]\right| \geq \varepsilon\right) \leq \frac{\sum_{j=1}^n \text{Var}[X_j]}{n^2 \varepsilon^2} = \frac{\text{Var}[X]}{n \varepsilon^2} \rightarrow 0$$

grâce à l'inégalité de Tchebychev. Dans le cas L^2 , on peut remplacer la convergence en probabilité par la convergence plus forte en norme L^2 .

Théorème 5.4.1 (Loi faible des grands nombres, cas L^2) Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de loi commune $X \in L^2$. Alors

$$\bar{X}^{(n)} \xrightarrow{\text{qm}} \mathbb{E}[X].$$

Démonstration : Notons $\mu = \mathbb{E}[X]$. Un calcul direct montre que

$$\mathbb{E} \left[\left| \bar{X}^{(n)} - \mu \right|^2 \right] = \mathbb{E} \left[\left| \frac{1}{n} \sum_{j=1}^n (X_j - \mu) \right|^2 \right] = \frac{1}{n^2} \sum_{j=1}^n \mathbb{E}[(X_j - \mu)^2] = \frac{\text{Var}[X]}{n},$$

ce qui nous apporte le résultat. ■

Remarque 5.4.2 Remarquons que dans la preuve précédente il aurait suffi de demander

$$\frac{1}{n^2} \sum_{j=1}^n \mathbb{E}[(X_j - \mu)^2] \rightarrow 0,$$

donc l'hypothèse d'identique distribution n'est pas indispensable.

Intéressons-nous à présent à l'hypothèse plus faible $X \in L^1$.

Théorème 5.4.3 (Loi faible des grands nombres, cas L^1 , Khintchine) Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de loi commune $X \in L^1$. Alors

$$\bar{X}^{(n)} \xrightarrow{\mathbb{P}} \mathbb{E}[X].$$

Démonstration : Quitte à remplacer X_n par $X_n - \mathbb{E}[X]$ pour tout n , on peut supposer que $\mathbb{E}[X] = 0$. Remarquons tout d'abord que puisque $\mathbf{1}_{\{|X| \geq M\}}|X| \leq |X|$ et $X \in L^1$, le théorème de la convergence dominée implique que

$$\lim_{M \rightarrow +\infty} \mathbb{E}[\mathbf{1}_{\{|X| \geq M\}}|X|] = 0.$$

Fixons $\varepsilon_0 > 0$. Pour tout $\varepsilon \in]0, \frac{\varepsilon_0}{2}[$, il existe donc $M > 0$ tel que

$$\mathbb{E}[\mathbf{1}_{\{|X| \geq M\}}|X|] \leq \varepsilon^2.$$

Cette relation est vérifiée pour tout X_n puisque les variables aléatoires X_n ont la même loi que X . Posons

$$Y_n = \mathbf{1}_{\{|X_n| < M\}}X_n \quad \text{et} \quad Z_n = X_n - Y_n = \mathbf{1}_{\{|X_n| \geq M\}}X_n$$

de sorte que

$$\bar{X}^{(n)} = \bar{Y}^{(n)} - \mathbb{E}[\bar{X}^{(n)}] = (\bar{Y}^{(n)} - \mathbb{E}[\bar{Y}^{(n)}]) + (\bar{Z}^{(n)} - \mathbb{E}[\bar{Z}^{(n)}]).$$

Pour le premier terme, on applique l'inégalité de Tchebychev donnée dans le Corollaire 3.5.4 puisque les variables aléatoires Y_n sont bornées et donc appartiennent à L^2 . On a ainsi

$$\mathbb{P} \left(|\bar{Y}^{(n)} - \mathbb{E}[\bar{Y}^{(n)}]| \geq \varepsilon \right) \leq \frac{\sum_{j=1}^n \text{Var}[Y_j]}{\varepsilon^2 n^2} \leq \frac{\sum_{j=1}^n \mathbb{E}[Y_j^2]}{\varepsilon^2 n^2} \leq \frac{M^2}{\varepsilon^2 n}$$

qui tend vers 0 lorsque n tend vers l'infini. Ainsi, il existe $M' \geq M$ tel que

$$\mathbb{P}\left(|\bar{Y}^{(n)} - \mathbb{E}[\bar{Y}^{(n)}]| \geq \varepsilon\right) \leq \varepsilon$$

si $n \geq M'$.

Pour le deuxième terme, on utilise l'inégalité de Markov pour écrire

$$\begin{aligned} \mathbb{P}\left(|\bar{Z}^{(n)} - \mathbb{E}[\bar{Z}^{(n)}]| \geq \varepsilon\right) &\leq \frac{\mathbb{E}[|\bar{Z}^{(n)} - \mathbb{E}[\bar{Z}^{(n)}]|]}{\varepsilon} \\ &\leq \frac{\mathbb{E}[|\bar{Z}^{(n)}|] + |\mathbb{E}[\bar{Z}^{(n)}]|}{\varepsilon} \\ &\leq \frac{2}{n\varepsilon} \sum_{j=1}^n \mathbb{E}[|Z_j|] \\ &= \frac{2}{n\varepsilon} \sum_{j=1}^n \mathbb{E}[\mathbf{1}_{\{|X_j| \geq M\}} |X_j|] \\ &\leq 2\varepsilon \end{aligned}$$

vu le choix de M .

Au total, on obtient que

$$\begin{aligned} \mathbb{P}(|\bar{X}^{(n)} - \mathbb{E}[\bar{X}^{(n)}]| \geq \varepsilon_0) &\leq \mathbb{P}(|\bar{Y}^{(n)} - \mathbb{E}[\bar{Y}^{(n)}]| \geq \frac{\varepsilon_0}{2}) + \mathbb{P}(|\bar{Z}^{(n)} - \mathbb{E}[\bar{Z}^{(n)}]| \geq \frac{\varepsilon_0}{2}) \\ &\leq \mathbb{P}(|\bar{Y}^{(n)} - \mathbb{E}[\bar{Y}^{(n)}]| \geq \varepsilon) + \mathbb{P}(|\bar{Z}^{(n)} - \mathbb{E}[\bar{Z}^{(n)}]| \geq \varepsilon) \\ &\leq 3\varepsilon \end{aligned}$$

pour tout $n \geq M'$. On conclut en prenant ε arbitrairement petit. ■

Remarque 5.4.4 L'hypothèse $X \in L^1$ est optimale puisqu'elle est nécessaire pour que la limite soit bien définie et finie. On peut montrer que la convergence a lieu également dans L^1 : nous ne le ferons pas ici, on peut l'obtenir comme conséquence des résultats de la théorie des martingales.

La loi faible des grands nombres ne concerne que des types de convergence "faibles". Il est donc intéressant de se demander si on peut remplacer ces convergences par la convergence "forte", c'est-à-dire la convergence presque sûre. Pour cela, nous allons tout d'abord nous intéresser à la convergence presque sûre de séries de variables aléatoires indépendantes. La loi forte des grands nombres s'en déduira alors via l'utilisation d'un résultat d'analyse sur la convergence des séries.

On sait, vu la loi du 0-1 de Kolmogorov que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes, alors l'événement $\{\sum_n X_n \text{ existe}\}$ est de probabilité 0 ou 1. Un critère de convergence va être obtenu comme conséquence de la célèbre inégalité suivante déjà rencontrée dans le chapitre précédent. Une preuve alternative n'utilisant pas la théorie des martingales est proposée ci-dessous

Théorème 5.4.5 (Inégalité de Kolmogorov) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléa-

toires indépendantes et centrées de L^2 . Pour tout $a > 0$, on a

$$\mathbb{P} \left(\sup_{n \in \mathbb{N}} \left| \sum_{k=0}^n X_k \right| > a \right) \leq \frac{1}{a^2} \sum_{k=0}^{+\infty} \mathbb{E}[X_k^2].$$

Démonstration : Bien sûr, on peut supposer que $\sum_{k=0}^{+\infty} \mathbb{E}[X_k^2] < +\infty$. Alors, la suite des sommes partielles des $(X_n)_{n \in \mathbb{N}}$ est de Cauchy dans L^2 puisque par indépendance

$$\mathbb{E} \left[\left(\sum_{k=p}^q X_k \right)^2 \right] = \sum_{k=p}^q \mathbb{E}[X_k^2] \rightarrow 0$$

lorsque p, q tendent vers l'infini. Puisque l'espace L^2 est de Banach, cela implique que la série $\sum_{k=0}^{+\infty} X_k$ converge dans L^2 et est donc presque sûrement finie.

Considérons à présent la variable aléatoire T_a à valeurs dans $\mathbb{N} \cup \{+\infty\}$ définie par

$$T_a = \begin{cases} +\infty & \text{si } \left| \sum_{k=0}^n X_k \right| \leq a \text{ pour tout } n \in \mathbb{N}, \\ \min \left\{ n \in \mathbb{N} : \left| \sum_{k=0}^n X_k \right| > a \right\} & \text{sinon.} \end{cases}$$

Ainsi, T_a est le premier indice pour lequel la somme va dépasser a . Clairement, on a

$$\{T_a \leq m\} = \bigcup_{n \leq m} \left\{ \left| \sum_{k=0}^n X_k \right| > a \right\} \tag{5.1}$$

et il s'ensuit que

$$\{T_a < +\infty\} = \bigcup_{m \in \mathbb{N}} \{T_a \leq m\} = \bigcup_{n \in \mathbb{N}} \left\{ \left| \sum_{k=0}^n X_k \right| > a \right\} = \left\{ \sup_{n \in \mathbb{N}} \left| \sum_{k=0}^n X_k \right| > a \right\}.$$

Par conséquent, il suffit de montrer que

$$\mathbb{P}(T_a < +\infty) \leq \frac{1}{a^2} \sum_{k=0}^{+\infty} \mathbb{E}[X_k^2].$$

Pour cela, définissons une suite $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires via la relation

$$Y_n = X_n \mathbb{1}_{\{T_a \geq n\}}.$$

Comme $|Y_n| \leq |X_n|$, on a $Y_n \in L^2$. De plus, on a

$$\sum_{k=0}^{T_a} X_k = \sum_{k=0}^{+\infty} Y_k$$

et donc

$$\{T_a < +\infty\} \subseteq \left\{ \left| \sum_{k=0}^{T_a} X_k \right| > a \right\} = \left\{ \left| \sum_{k=0}^{+\infty} Y_k \right| > a \right\}.$$

Notons que le même raisonnement que celui fait pour la suite $(X_n)_{n \in \mathbb{N}}$ montre que la série $\sum_{k=0}^{+\infty} Y_k$ converge dans L^2 et donc presque sûrement vers une variable aléatoire finie. Si on montre que $\mathbb{E}[Y_n Y_m] = 0$ pour tout $n \neq m$, alors l'inégalité de Markov appliqué à la variable aléatoire $\sum_{k=0}^{+\infty} Y_k$ implique que

$$\begin{aligned} \mathbb{P}(T_a < +\infty) &\leq \mathbb{P}\left(\left|\sum_{k=0}^{+\infty} Y_k\right| > a\right) \\ &\leq \frac{1}{a^2} \mathbb{E}\left[\left|\sum_{k=0}^{+\infty} Y_k\right|^2\right] \\ &= \frac{1}{a^2} \sum_{k=0}^{+\infty} \mathbb{E}[|Y_k|^2] \\ &\leq \frac{1}{a^2} \sum_{k=0}^{+\infty} \mathbb{E}[|X_k|^2], \end{aligned}$$

ce qui donnera la conclusion.

Pour montrer que $\mathbb{E}[Y_n Y_m] = 0$, on remarque que l'égalité (5.1) implique que pour tout $m \in \mathbb{N}$, l'événement $\{T_a \leq m\}$ est dans la σ -algèbre $\sigma(X_0, \dots, X_m)$ engendrée par X_0, \dots, X_m . Par conséquent, l'événement $\{T_a \geq m\} = \{T_a \leq m-1\}^c$ est dans la σ -algèbre $\sigma(X_0, \dots, X_{m-1})$ engendrée par X_0, \dots, X_{m-1} . On en tire que pour tout $n \leq m-1$, la variable aléatoire $X_n \mathbb{1}_{\{T_a \geq m\}}$ est $\sigma(X_0, \dots, X_{m-1})$ mesurable et donc $\sigma(X_n \mathbb{1}_{\{T_a \geq m\}}) \subseteq \sigma(X_0, \dots, X_{m-1})$. Par le Théorème 3.2.4 d'indépendance par paquets, les σ -algèbres $\sigma(X_0, \dots, X_{m-1})$ et $\sigma(X_m)$ sont indépendantes, et donc il en est de même pour $\sigma(X_n \mathbb{1}_{\{T_a \geq m\}})$ et $\sigma(X_m)$. Cela signifie que la variable aléatoire $X_n \mathbb{1}_{\{T_a \geq m\}}$ est indépendante de X_m et donc

$$\mathbb{E}[X_n \mathbb{1}_{\{T_a \geq m\}} X_m] = \mathbb{E}[X_n \mathbb{1}_{\{T_a \geq m\}}] \mathbb{E}[X_m] = 0.$$

Or, on a

$$X_n \mathbb{1}_{\{T_a \geq m\}} X_m = X_n \mathbb{1}_{\{T_a \geq n\}} X_m \mathbb{1}_{\{T_a \geq m\}} = Y_n Y_m$$

d'où $\mathbb{E}[Y_n Y_m] = 0$. ■

Remarque 5.4.6 Par l'inégalité de Tchebychev, on sait que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires centrées de L^2 , alors

$$\mathbb{P}\left(\left|\sum_{k=0}^n X_k\right| \geq a\right) \leq \frac{1}{a^2} \sum_{k=0}^n \mathbb{E}[X_k^2] \leq \frac{1}{a^2} \sum_{k=0}^{+\infty} \mathbb{E}[X_k^2].$$

Ainsi, l'inégalité de Kolmogorov renforce l'inégalité de Tchebychev de la même sorte que la convergence presque sûre renforce la convergence en probabilité comme vu dans le Lemme 5.3.5.

Théorème 5.4.7 (Kolmogorov) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et centrées de L^2 . Si

$$\sum_{n \in \mathbb{N}} \text{Var}[X_n] = \sum_{n \in \mathbb{N}} \mathbb{E}[X_n^2] < +\infty,$$

alors la série

$$\sum_{n \in \mathbb{N}} X_n$$

converge presque sûrement vers une variable aléatoire finie.

Démonstration : Posons pour tout $n \in \mathbb{N}$

$$S_n = \sum_{k=0}^n X_k.$$

Remarquons que

$$\limsup_{n \rightarrow +\infty} S_n - \liminf_{n \rightarrow +\infty} S_n = \lim_{m \rightarrow +\infty} \left(\sup_{n > m} S_n - \inf_{n > m} S_n \right) = \inf_{m \in \mathbb{N}} \left(\sup_{n > m} S_n - \inf_{n > m} S_n \right)$$

et donc pour tout $\varepsilon > 0$, on a

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} S_n - \liminf_{n \rightarrow +\infty} S_n > \varepsilon \right) \leq \mathbb{P} \left(\sup_{n > m} S_n - \inf_{n > m} S_n > \varepsilon \right) \quad (5.2)$$

pour tout $m \in \mathbb{N}$. De plus, on a

$$S_{n_1} - S_{n_2} \leq |S_{n_1} - S_m| + |S_{n_2} - S_m|$$

de sorte que

$$\sup_{n > m} S_n - \inf_{n > m} S_n \leq 2 \sup_{n > m} |S_n - S_m|.$$

Ainsi, en appliquant l'inégalité de Kolmogorov 5.4.5 à la suite $(X_n)_{n \geq m+1}$, on obtient

$$\begin{aligned} \mathbb{P} \left(\sup_{n > m} S_n - \inf_{n > m} S_n > \varepsilon \right) &\leq \mathbb{P} \left(\sup_{n > m} |S_n - S_m| > \varepsilon/2 \right) \\ &= \mathbb{P} \left(\sup_{n \geq m+1} \left| \sum_{k=m+1}^n X_k \right| > \varepsilon/2 \right) \\ &\leq \frac{4}{\varepsilon^2} \sum_{k=m+1}^{+\infty} \mathbb{E}[X_k^2] \end{aligned}$$

qui par hypothèse tend vers 0 lorsque m tend vers l'infini. Par conséquent, l'inégalité (5.2) implique que

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} S_n - \liminf_{n \rightarrow +\infty} S_n > \varepsilon \right) = 0.$$

Remarquons à présent que

$$\left\{ \limsup_{n \rightarrow +\infty} S_n \neq \liminf_{n \rightarrow +\infty} S_n \right\} = \bigcup_{p \in \mathbb{N}_0} \left\{ \limsup_{n \rightarrow +\infty} S_n - \liminf_{n \rightarrow +\infty} S_n > 1/p \right\}$$

et donc cet événement a une probabilité nulle. Il s'ensuit que presque sûrement, on a

$$\limsup_{n \rightarrow +\infty} S_n = \liminf_{n \rightarrow +\infty} S_n$$

et donc que $\lim_{n \rightarrow +\infty} S_n = \sum_{n \in \mathbb{N}} X_n$ existe. Pour conclure, il reste à s'assurer que la limite est finie. Or, on a la convergence de la série dans L^2 puisque, en utilisant le Lemme de Fatou et l'indépendance des variables aléatoires, on peut écrire

$$\begin{aligned} \mathbb{E} \left[\left| \sum_{k>n} X_k \right|^2 \right] &= \mathbb{E} \left[\liminf_{m \rightarrow +\infty} \left(\sum_{k=n+1}^m X_k \right)^2 \right] \\ &\leq \liminf_{m \rightarrow +\infty} \mathbb{E} \left[\left(\sum_{k=n+1}^m X_k \right)^2 \right] \\ &= \sum_{k=n+1}^{+\infty} \mathbb{E}[X_k^2] \end{aligned}$$

qui tend vers 0 par hypothèse lorsque n tend vers l'infini. Donc la série est presque sûrement finie. On peut aussi procéder en remarquant que la série est de Cauchy dans L^2 et donc converge dans L^2 vers une variable aléatoire presque sûrement finie. ■

Exemple 5.4.8 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes telles que

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2},$$

i.e. chaque variable X_n modélise un choix de signe uniforme. Si $\alpha > 1/2$, alors

$$\text{Var} \left[\frac{X_n}{n^\alpha} \right] = \frac{1}{n^{2\alpha}}$$

et donc la série

$$\sum_{n \in \mathbb{N}} \frac{X_n}{n^\alpha}$$

converge presque sûrement. Remarquons que pour que la série converge pour tout choix de signes, il faut imposer $\alpha > 1$.

Remarque 5.4.9 Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes, alors on peut appliquer le Théorème de Kolmogorov à la suite $(X_n - \mathbb{E}[X_n])_{n \in \mathbb{N}}$ et on obtient que si

$$\sum_{n \in \mathbb{N}} \text{Var}[X_n] < +\infty,$$

alors la série $\sum_{n \in \mathbb{N}} (X_n - \mathbb{E}[X_n])$ converge presque sûrement.

Pour démontrer la loi forte des grands nombres dans le cas L^2 , on peut utiliser le lemme d'analyse suivant.

Lemme 5.4.10 (Lemme de Kronecker) Si $(a_n)_{n \in \mathbb{N}_0}$ est une suite croissante de réels positifs qui converge vers l'infini, alors pour toute suite $(x_n)_{n \in \mathbb{N}}$ de réels telle que $\sum_{n \in \mathbb{N}_0} \frac{x_n}{a_n}$

converge dans \mathbb{R} , on a

$$\lim_{n \rightarrow +\infty} \frac{1}{a_n} \sum_{k=1}^n x_k = 0.$$

Démonstration : Posons $u_n = \sum_{j=n}^{+\infty} \frac{x_j}{a_j}$ pour tout $n \geq 1$. On remarque que l'on peut écrire successivement

$$\begin{aligned} \frac{1}{a_n} \sum_{k=1}^n x_k &= \frac{1}{a_n} \sum_{k=1}^n a_k \frac{x_k}{a_k} \\ &= \frac{1}{a_n} \sum_{k=1}^n a_k (u_k - u_{k+1}) \\ &= \frac{1}{a_n} \left(\sum_{k=2}^n (a_k - a_{k-1}) u_k + a_1 u_1 - a_n u_{n+1} \right) \\ &= \frac{1}{a_n} \sum_{k=2}^n (a_k - a_{k-1}) u_k + \frac{a_1 u_1}{a_n} - u_{n+1}. \end{aligned}$$

Par hypothèse, les deux derniers termes tendent vers 0 lorsque n tend vers l'infini et il reste à traiter le premier terme. Soit $\varepsilon > 0$. Comme la suite $(u_n)_{n \in \mathbb{N}_0}$ tend vers 0, on peut trouver N tel que $|u_n| < \varepsilon$ pour tout $n \geq N$. Puisque N est fixé et que la suite $(a_n)_{n \in \mathbb{N}_0}$ tend vers l'infini, il existe N' tel que

$$\frac{1}{a_n} \sum_{k=2}^{N-1} (a_k - a_{k-1}) |u_k| < \varepsilon$$

pour tout $n \geq N'$. Il vient alors que pour tout $n \geq \max\{N, N'\}$,

$$\left| \frac{1}{a_n} \sum_{k=2}^n (a_k - a_{k-1}) u_k \right| \leq \frac{1}{a_n} \sum_{k=2}^{N-1} (a_k - a_{k-1}) |u_k| + \frac{1}{a_n} \sum_{k=N}^n (a_k - a_{k-1}) |u_k| < \varepsilon + \varepsilon \left(1 - \frac{a_{N-1}}{a_n} \right) < 2\varepsilon$$

puisque la suite $(a_n)_{n \in \mathbb{N}_0}$ est croissante. ■

Théorème 5.4.11 (Loi forte des grands nombres, cas L^2) Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de loi commune $X \in L^2$. Alors

$$\bar{X}^{(n)} \xrightarrow{p.s.} \mathbb{E}[X].$$

Démonstration : Sans perte de généralité, quitte à remplacer X_n par $X_n - \mathbb{E}[X_n]$, on peut supposer que $\mathbb{E}[X] = \mathbb{E}[X_n] = 0$ pour tout $n \in \mathbb{N}_0$. Remarquons que si $Y_n = \frac{X_n}{n}$, alors

$$\sum_{n \in \mathbb{N}_0} \mathbb{E}[Y_n^2] = \sum_{n \in \mathbb{N}_0} \frac{\mathbb{E}[X^2]}{n^2} < +\infty$$

et le Théorème de Kolmogorov implique donc que la série

$$\sum_{n \in \mathbb{N}_0} Y_n = \sum_{n \in \mathbb{N}_0} \frac{X_n}{n}$$

converge presque sûrement vers une variable aléatoire finie. Le Lemme de Kronecker donne la conclusion. ■

La même preuve donne en fait un résultat plus fort qui garantit la convergence presque sûre sans demander que les variables soient identiquement distribuées.

Théorème 5.4.12 (Rajchman) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et centrées de L^2 et $(a_n)_{n \in \mathbb{N}}$ une suite croissante de réels positifs qui converge vers l'infini. Si

$$\sum_{n \in \mathbb{N}} \frac{\mathbb{E}[X_n^2]}{a_n^2} < +\infty,$$

alors

$$\frac{1}{a_n} \sum_{k=0}^n X_k \xrightarrow{p.s.} 0.$$

Il reste à présent à enlever l'hypothèse d'appartenance à L^2 pour obtenir le résultat le plus fort.

Théorème 5.4.13 (Loi forte des grands nombres, cas L^1) Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de loi commune $X \in L^1$. Alors

$$\bar{X}^{(n)} \xrightarrow{p.s.} \mathbb{E}[X].$$

Démonstration : Sans perte de généralité, quitte à remplacer X_n par $X_n - \mathbb{E}[X_n]$, on peut supposer que $\mathbb{E}[X] = \mathbb{E}[X_n] = 0$ pour tout $n \in \mathbb{N}_0$. On va procéder comme dans la preuve de la loi faible des grands nombres en considérant des variables aléatoires tronquées qui appartiendront alors à L^2 : pour tout $n \in \mathbb{N}_0$, on pose

$$Y_n = X_n \mathbf{1}_{\{|X_n| < n\}}.$$

Comme $|Y_n| \leq n$, on a bien que $Y_n \in L^2$. Par conséquent, les variables aléatoires $Y_n - \mathbb{E}[Y_n]$ sont des variables aléatoires indépendantes et centrées de L^2 . Puisque les variables aléatoires X_n , $n \in \mathbb{N}_0$, sont identiquement distribuées et de même loi que X , on a

$$\mathbb{E}[Y_n^2] = \mathbb{E}[X_n^2 \mathbf{1}_{\{|X_n| < n\}}] = \mathbb{E}[X^2 \mathbf{1}_{\{|X| < n\}}] = \sum_{k=1}^n \mathbb{E}[X^2 \mathbf{1}_{\{k-1 \leq |X| < k\}}].$$

Remarquons que pour tout $k \geq 2$, on a

$$\sum_{n=k}^{+\infty} \frac{1}{n^2} \leq \int_{k-1}^{+\infty} \frac{1}{x^2} dx = \frac{1}{k-1} \leq \frac{2}{k}$$

et donc il existe une constante $C > 0$ telle que

$$\sum_{n=k}^{+\infty} \frac{1}{n^2} \leq \frac{C}{k}$$

pour tout $k \in \mathbb{N}_0$. Par conséquent, on peut écrire

$$\begin{aligned}
 \sum_{n \in \mathbb{N}_0} \frac{\text{Var}[Y_n]}{n^2} &\leq \sum_{n \in \mathbb{N}_0} \frac{\mathbb{E}[Y_n^2]}{n^2} \\
 &= \sum_{n \in \mathbb{N}_0} \frac{\mathbb{E}[X_n^2 \mathbf{1}_{\{|X_n| < n\}}]}{n^2} \\
 &= \sum_{n \in \mathbb{N}_0} \frac{1}{n^2} \sum_{k=1}^n \mathbb{E}[X^2 \mathbf{1}_{\{k-1 \leq |X| < k\}}] \\
 &= \sum_{k \in \mathbb{N}_0} \mathbb{E}[X^2 \mathbf{1}_{\{k-1 \leq |X| < k\}}] \sum_{n=k}^{+\infty} \frac{1}{n^2} \\
 &\leq \sum_{k \in \mathbb{N}_0} \mathbb{E}[k|X| \mathbf{1}_{\{k-1 \leq |X| < k\}}] \frac{C}{k} \\
 &\leq C \sum_{k \in \mathbb{N}_0} \mathbb{E}[|X| \mathbf{1}_{\{k-1 \leq |X| < k\}}] \\
 &= C \mathbb{E}[|X|] < +\infty
 \end{aligned}$$

où on peut échanger les sommes car tout est positif et en utilisant le théorème de la convergence monotone puisque $|X| = \sum_{k \in \mathbb{N}_0} |X| \mathbf{1}_{\{k-1 \leq |X| < k\}}$. Par le Théorème de Rajchman, on obtient que

$$\frac{1}{n} \sum_{k=1}^n (Y_k - \mathbb{E}[Y_k]) \xrightarrow{\text{p.s.}} 0.$$

Or, le théorème de la convergence dominée implique que

$$\mathbb{E}[Y_n] = \mathbb{E}[X_n \mathbf{1}_{\{|X_n| < n\}}] = \mathbb{E}[X \mathbf{1}_{\{|X| < n\}}] \rightarrow 0 \quad \text{si } n \rightarrow +\infty$$

et donc les moyennes de Césaro convergent également vers 0, c'est-à-dire

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E}[Y_k] \rightarrow 0 \quad \text{si } n \rightarrow +\infty.$$

On a donc

$$\frac{1}{n} \sum_{k=1}^n Y_k = \frac{1}{n} \sum_{k=1}^n (Y_k - \mathbb{E}[Y_k]) + \frac{1}{n} \sum_{k=1}^n \mathbb{E}[Y_k] \xrightarrow{\text{p.s.}} 0.$$

Pour conclure, il suffit de montrer que

$$\frac{1}{n} \sum_{k=1}^n (X_n - Y_n) \xrightarrow{\text{p.s.}} 0,$$

ce qui sera à nouveau garanti par les propriétés des moyennes de Césaro si $(X_n - Y_n) \xrightarrow{\text{p.s.}} 0$, avec $X_n - Y_n = X_n \mathbf{1}_{\{|X_n| \geq n\}}$. Il suffit d'appliquer la Proposition 5.1.5 puisque par (2.2)

$$\sum_{n \in \mathbb{N}_0} \mathbb{P}(|X_n| \mathbf{1}_{\{|X_n| \geq n\}} > \varepsilon) = \sum_{n \in \mathbb{N}_0} \mathbb{P}(|X_n| \geq n) = \sum_{n \in \mathbb{N}_0} \mathbb{P}(|X| \geq n) < +\infty$$

puisque $X \in L^1$. ■

Présentons quelques applications de la loi des grands nombres.

Exemple 5.4.14 (Approche fréquentiste) Soit A un événement dont on souhaite estimer la probabilité. Remarquons que $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$ et donc estimer la probabilité de A revient à estimer l'espérance d'une variable aléatoire. Ainsi, on note $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires i.i.d. de même loi que $\mathbb{1}_A$. Alors

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{p.s.}} \mathbb{P}(A).$$

Ce résultat fait le lien entre l'approche axiomatique moderne des probabilités et l'approche fréquentiste historique.

Exemple 5.4.15 (Théorème de Glivenko-Cantelli) La loi des grands nombres permet de démontrer la convergence ponctuelle presque sûre de la fonction de répartition empirique d'un échantillon vers la fonction de répartition théorique. Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de fonction de répartition commune F . Posons

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k \leq x\}}$$

qui est la fonction de répartition empirique associée à l'échantillon X_1, \dots, X_n . On peut écrire $F(x) = \mathbb{E}[\mathbb{1}_{\{X \leq x\}}]$ où $X \sim F$. Alors, par la loi forte des grands nombres,

$$F_n(x) \xrightarrow{\text{p.s.}} F(x)$$

pour tout x . On peut montrer également que cette convergence est presque sûrement uniforme, c'est-à-dire

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\text{p.s.}} 0$$

presque sûrement. Ce résultat est connu sous le nom de théorème de Glivenko-Cantelli.

Exemple 5.4.16 Soient f une fonction continue sur $[0, 1]$ et

$$f_n(x) = \sum_{m=0}^n C_n^m x^m (1-x)^{n-m} f(m/n)$$

le polynôme de Bernstein de degré n associé. Alors

$$\sup_{x \in [0,1]} |f_n(x) - f(x)| \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

En effet, commençons par remarquer que $f_n(x) = \mathbb{E}[f(S_n/n)]$ avec $S_n = \sum_{k=1}^n X_k$ où les variables aléatoires X_k sont i.i.d. de loi de Bernoulli de paramètre x . Or f étant continue sur $[0, 1]$, elle est bornée par une constante B et uniformément continue. On en déduit que pour tout $\varepsilon > 0$ il existe un $\delta > 0$ tel que

$$|\mathbb{E}[f(S_n/n) - f(x)]| \leq \varepsilon + 2B\mathbb{P}[|S_n/n - x| > \delta].$$

Mais, par la loi faible des grands nombres, on sait que $\frac{S_n}{n} \xrightarrow{\mathbb{P}} x$. L'inégalité de Tchebychev montre en fait que la convergence est uniforme en x puisque $\text{Var}[S_n] = nx(1-x) \leq n$. On en tire donc finalement que $\lim_{n \rightarrow +\infty} |\mathbb{E}[f(S_n/n) - f(x)]| \leq \varepsilon$ pour tout $\varepsilon > 0$.

Exemple 5.4.17 Soient $(z_n)_{n \in \mathbb{N}_0}$ une suite de réels et $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires indépendantes de variance commune σ^2 satisfaisant la régression linéaire $\mathbb{E}[X_n] = \alpha + \beta z_n$ pour tout $n \in \mathbb{N}_0$. Les estimateurs des moindres carrés de α et β basé sur X_1, \dots, X_n sont donnés par $\hat{\beta}_n = \sum_{i=1}^n X_i(z_i - \bar{z}_n) / \sum_{i=1}^n (z_i - \bar{z}_n)^2$ et $\hat{\alpha}_n = \bar{X}_n - \hat{\beta}_n \bar{z}_n$, où $\bar{z}_n = (1/n) \sum_{i=1}^n z_i$. On montre que $\hat{\beta}_n$ est un estimateur sans biais qui satisfait $\text{Var}[\hat{\beta}_n] = \sigma^2 / \sum_{j=1}^n (z_j - \bar{z}_n)^2$ donc $\hat{\beta}_n \xrightarrow{\text{qm}} \beta$ si et seulement si $\sum_{j=1}^n (z_j - \bar{z}_n)^2 \rightarrow +\infty$. Pour les mêmes raisons, $\hat{\alpha}_n \xrightarrow{\text{qm}} \alpha$ si et seulement si $\bar{z}_n^2 / \sum_{j=1}^n (z_j - \bar{z}_n)^2 \rightarrow 0$.

Exemple 5.4.18 (Méthodes de Monte-Carlo) Soit B un ensemble borélien borné de \mathbb{R} . Si $f : B \rightarrow \mathbb{R}$ est intégrable et si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes et distribuées uniformément sur B , on a

$$\frac{1}{n} \sum_{j=1}^n f(X_j) \xrightarrow{\text{p.s.}} \mathbb{E}[f(X_1)] = \frac{1}{\lambda(B)} \int_B f(x) dx$$

par la loi forte des grands nombres. Les méthodes de Monte-Carlo sont basées sur cette convergence et permettent de faire des calculs approchés d'intégrales via la loi des grands nombres.

5.5 Convergence en loi et théorème central limite

Nous venons de voir que si $(X_n)_{n \in \mathbb{N}_0}$ est une suite de variables aléatoires indépendantes et de même loi $X \in L^1$, alors

$$\bar{X}^{(n)} = \frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\text{p.s.}} \mathbb{E}[X].$$

Cela nous apprend qu'un estimateur de $\mathbb{E}[X]$ est donné par $\bar{X}^{(n)}$. Cette méthode est pertinente car $\mathbb{E}[\bar{X}^{(n)}] = \mathbb{E}[X]$ (l'estimateur est non-biaisé) et si $n \rightarrow +\infty$, l'estimation se fait finalement sans erreur. Par conséquent, si on répète une infinité de fois l'expérience, on pourra calculer sans se tromper $\mathbb{E}[X]$. Évidemment, on ne pourra jamais prendre $n \rightarrow +\infty$ et en pratique, on travaillera avec un nombre n d'expériences fixé. Une erreur sera commise dans l'estimation de $\mathbb{E}[X]$ par $\bar{X}^{(n)}$. Lorsque n est grand, il est donc légitime de s'interroger sur l'ordre de grandeur de $\bar{X}^{(n)} - \mathbb{E}[X]$. La difficulté vient du fait que la loi de $\bar{X}^{(n)} - \mathbb{E}[X]$ dépend a priori de la distribution de X , qui en pratique est très souvent inconnue.

Exemple 5.5.1

- Si X_1, \dots, X_n sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, alors $X_1 + X_2 + \dots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$, et donc on obtient que $\bar{X}^{(n)} \sim \mathcal{N}(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}) = \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- Si X_1, \dots, X_n sont i.i.d. $\text{Bern}(p) = \text{Bin}(1, p)$, alors $\bar{X}^{(n)} \sim \frac{1}{n} \text{Bin}(n, p)$.
- Si X_1, \dots, X_n sont i.i.d. $\text{Pois}(\lambda)$, alors $\bar{X}^{(n)} \sim \frac{1}{n} \text{Pois}(n\lambda)$.

Le Théorème Central Limite, qui est l'un des résultats les plus importants de la théorie des probabilités, montre que quelle que soit la nature de la distribution de X , dès que celle-ci admet une variance, une somme de variables aléatoires i.i.d. de même loi que X converge toujours vers une loi gaussienne. Pour cela, nous allons introduire une notion de convergence plus faible : la convergence en loi.

Définition 5.5.2 Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire X si

$$\lim_{n \rightarrow +\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)] \quad \text{pour toute fonction } h \in C_b(\mathbb{R})$$

où $C_b(\mathbb{R})$ désigne l'ensemble des fonctions continues et bornées sur \mathbb{R} . Dans ce cas, on écrit

$$X_n \xrightarrow{\mathcal{L}} X.$$

Notons que si $h \in C_b(\mathbb{R})$, alors les espérances ci-dessous ont bien du sens. Remarquons également que la condition se réécrit

$$\int_{\mathbb{R}} h(x) d\mathbb{P}_{X_n}(x) \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x) \quad \forall h \in C_b(\mathbb{R}),$$

ce qui montre que cette notion de convergence ne fait intervenir que la loi des variables aléatoires, et pas les variables aléatoires elles-mêmes. Cette notion de convergence est beaucoup plus faible que celles de la section précédente : on n'exige pas que les variables donnent des valeurs proches, on n'exige même pas qu'elles soient définies sur un même espace probabilisé. En particulier, la convergence en loi n'implique aucun autre type de convergence. Nous verrons néanmoins un cas particulier où la convergence en loi d'une suite de variables aléatoires définies sur le même espace probabilisé implique la convergence en probabilité.

Remarque 5.5.3 La convergence en loi est très différente des autres types de convergence. Notons par exemple les deux faits suivants.

- Si $X_n \xrightarrow{\mathcal{L}} X$ et $X_n \xrightarrow{\mathcal{L}} Y$, alors on n'a pas nécessairement que $X = Y$ presque sûrement. Il suffit pour cela de prendre deux variables aléatoires X et Y de même loi mais différentes.
- Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ et si X_n et Y_n sont définies sur le même espace probabilisé, on ne peut rien dire sur la convergence de la somme et du produit. En effet, si on pose $X_n = X \sim \mathcal{N}(0, 1) \xrightarrow{\mathcal{L}} X$ et $Y_n = X \xrightarrow{\mathcal{L}} -X$ puisque $X \stackrel{\mathcal{L}}{=} -X$. Alors $X_n + Y_n = 2X \neq 0$ et $X_n Y_n = X^2 \neq -X^2$.

Il y a en fait un abus de langage à dire qu'une suite de variables aléatoires converge en loi vers X puisque la variable aléatoire X n'est pas définie de manière unique : seule sa loi \mathbb{P}_X l'est (ce qui est évident puisque la fonction caractéristique caractérise la loi et que les fonctions $x \mapsto e^{itx}$ sont continues sur \mathbb{R} et bornées). On devrait en fait de parler de convergence pour les lois et donc comprendre que la suite $(\mathbb{P}_{X_n})_{n \in \mathbb{N}}$ converge vers \mathbb{P}_X . Signalons que l'espace des mesures de probabilité sur \mathbb{R} peut être vu comme un sous-ensemble du dual (formé des formes linéaires continues) de l'espace $C_b(\mathbb{R})$ muni de la norme $\|h\|_{\infty} = \sup_{x \in \mathbb{R}} |h(x)|$. La convergence en loi correspond alors à la convergence faible-* sur le dual (topologie de la convergence simple).

Avant de donner des exemples, démontrons des caractérisations équivalentes de la convergence en loi.

Proposition 5.5.4 Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et X une variable aléatoire. Les assertions suivantes sont équivalentes :

1. $X_n \xrightarrow{\mathcal{L}} X$,

2. $\lim_{n \rightarrow +\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)]$ pour toute fonction $h \in C_c(\mathbb{R})$, où $C_c(\mathbb{R})$ est l'ensemble des fonctions continues sur \mathbb{R} et à support compact,
3. $\lim_{n \rightarrow +\infty} \varphi^{X_n}(t) = \varphi^X(t)$ pour tout $t \in \mathbb{R}$,
4. $\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$ en tout point de continuité t de F_X (i.e. en tout point $t \in \mathbb{R}$ tel que $\mathbb{P}(X = t) = 0$).

Démonstration : $\underline{1} \Leftrightarrow \underline{2}$. Il est clair que la convergence en loi implique la deuxième condition puisque $C_c(\mathbb{R}) \subseteq C_b(\mathbb{R})$. Montrons l'implication inverse. Soit $h \in C_b(\mathbb{R})$. Considérons une suite croissante $(g_k)_{k \in \mathbb{N}}$ de fonctions continues à support compact et à valeurs dans $[0, 1]$ qui converge vers la fonction identiquement égale à 1. Remarquons à présent que

$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| &\leq |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X_n)g_k(X_n)]| + |\mathbb{E}[h(X_n)g_k(X_n)] - \mathbb{E}[h(X)g_k(X)]| \\ &\quad + |\mathbb{E}[h(X)g_k(X)] - \mathbb{E}[h(X)]|. \end{aligned}$$

Remarquons que le deuxième terme tend vers 0 lorsque n tend vers l'infini puisque pour tout $k \in \mathbb{N}$, on a $hg_k \in C_c(\mathbb{R})$. Le premier et le troisième termes se traitent de manière similaire. Pour le premier, on a

$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X_n)g_k(X_n)]| = \mathbb{E}[h(X_n)(1 - g_k(X_n))] \leq \sup_{x \in \mathbb{R}} |h(x)|(1 - \mathbb{E}[g_k(X_n)]).$$

Par hypothèse, $1 - \mathbb{E}[g_k(X_n)]$ tend vers $1 - \mathbb{E}[g_k(X)]$ lorsque n tend vers l'infini. Enfin, le troisième terme est majoré par

$$|\mathbb{E}[h(X)g_k(X)] - \mathbb{E}[h(X)]| \leq \sup_{x \in \mathbb{R}} |h(x)|(1 - \mathbb{E}[g_k(X)]).$$

Au total, on a donc

$$\limsup_{n \rightarrow +\infty} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq 2 \sup_{x \in \mathbb{R}} |h(x)|(1 - \mathbb{E}[g_k(X)])$$

quel que soit $k \in \mathbb{N}$. On conclut en prenant la limite lorsque k tend vers l'infini et en utilisant le théorème de la convergence monotone pour montrer que $1 - \mathbb{E}[g_k(X)]$ tend vers 0.

$\underline{1} \Rightarrow \underline{3}$. C'est évident puisque la fonction $x \mapsto e^{itx}$ appartient à $C_b(\mathbb{R})$ pour tout $t \in \mathbb{R}$.

$\underline{3} \Rightarrow \underline{2}^1$. Soit $h \in C_c(\mathbb{R})$. Comme h est continue et s'annule en dehors d'un compact, elle est bornée et on peut considérer $B > 0$ tel que $\sup_{x \in \mathbb{R}} |h(x)| \leq B$. Fixons $\varepsilon > 0$. Puisque h est nécessairement uniformément continue, il existe $\delta > 0$ tel que

$$|h(x) - h(y)| < \varepsilon \quad \forall x, y \text{ tels que } |x - y| < \delta.$$

Considérons une variable aléatoire $Y \sim \mathcal{N}(0, \sigma^2)$ indépendante de tout le reste avec $\sigma > 0$ choisi tel que $2B\mathbb{P}(|Y| \geq \delta) < \varepsilon^2$. Alors, on peut écrire

$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| &\leq |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X_n + Y)]| + |\mathbb{E}[h(X_n + Y)] - \mathbb{E}[h(X + Y)]| \\ &\quad + |\mathbb{E}[h(X + Y)] - \mathbb{E}[h(X)]|. \end{aligned}$$

1. La preuve proposée utilise l'hypothèse supplémentaire que les variables aléatoires sont définies sur le même espace probabilisé.

2. On peut le faire quitte à diminuer σ^2 par l'inégalité de Markov.

Les premier et troisième termes se résolvent de la même façon. Pour le premier, on a

$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X_n + Y)]| &\leq \mathbb{E}[|h(X_n) - h(X_n + Y)|] \\ &= \mathbb{E}[|h(X_n) - h(X_n + Y)| \mathbf{1}_{\{|Y| < \delta\}}] + \mathbb{E}[|h(X_n) - h(X_n + Y)| \mathbf{1}_{\{|Y| \geq \delta\}}] \\ &\leq \varepsilon + 2B\mathbb{P}(|Y| \geq \delta) \leq 2\varepsilon \end{aligned}$$

où la dernière ligne découle de la continuité uniforme de h et du choix de σ . On procède de la même manière pour le troisième terme et on montre donc que

$$|\mathbb{E}[h(X + Y)] - \mathbb{E}[h(X)]| \leq 2\varepsilon.$$

Il reste donc à étudier le deuxième terme. Vu l'indépendance entre X_n et Y , on a par Tonelli-Fubini (puisque $h(X_n + Y) \in L^1$)

$$\begin{aligned} \mathbb{E}[h(X_n + Y)] &= \iint h(x + y) d\mathbb{P}_Y(y) d\mathbb{P}_{X_n}(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int \left(\int_{\mathbb{R}} h(x + y) e^{-y^2/2\sigma^2} dy \right) d\mathbb{P}_{X_n}(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int \left(\int_{\mathbb{R}} h(u) e^{-(u-x)^2/2\sigma^2} du \right) d\mathbb{P}_{X_n}(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} h(u) \left(\int e^{-(u-x)^2/2\sigma^2} d\mathbb{P}_{X_n}(x) \right) du. \end{aligned}$$

Or, rappelons que si $W \sim \mathcal{N}(0, v^2)$, alors

$$\varphi^W(t) = e^{-v^2 t^2/2}$$

et donc on a pour $v = 1/\sigma$,

$$e^{-(u-x)^2/2\sigma^2} = \varphi^W(u-x) = \mathbb{E}[e^{i(u-x)W}] = \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i(u-x)w} e^{-w^2\sigma^2/2} dw.$$

Il s'ensuit

$$\begin{aligned} \mathbb{E}[h(X_n + Y)] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} h(u) \left(\int \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i(u-x)w} e^{-w^2\sigma^2/2} dw d\mathbb{P}_{X_n}(x) \right) du \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} h(u) \left(\int_{\mathbb{R}} e^{iuw} e^{-w^2\sigma^2/2} \int_{\mathbb{R}} e^{-ixw} d\mathbb{P}_{X_n}(x) dw \right) du \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} h(u) \left(\int_{\mathbb{R}} e^{iuw} e^{-w^2\sigma^2/2} \varphi^{X_n}(-w) dw \right) du \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-w^2\sigma^2/2} \varphi^{X_n}(-w) \left(\int_{\mathbb{R}} h(u) e^{iuw} du \right) dw. \end{aligned}$$

En utilisant le théorème de la convergence dominée en majorant la fonction caractéristique de X_n par 1, on déduit

$$\lim_{n \rightarrow +\infty} \mathbb{E}[h(X_n + Y)] = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-w^2\sigma^2/2} \varphi^X(-w) \left(\int_{\mathbb{R}} h(u) e^{iuw} du \right) dw.$$

En effectuant le même raisonnement que précédemment pour X à la place de X_n , on trouve que

$$\frac{1}{2\pi} \int_{\mathbb{R}} e^{-w^2\sigma^2/2} \varphi^X(-w) \left(\int_{\mathbb{R}} h(u) e^{iuw} du \right) dw = \mathbb{E}[h(X + Y)]$$

et donc le deuxième terme tend vers 0. Cela conclut cette partie de la preuve.

1 \Rightarrow 4. Supposons tout d'abord que $X_n \xrightarrow{\mathcal{L}} X$ et fixons un point t de continuité de F_X . Soit $(h_k)_{k \in \mathbb{N}_0}$ la suite de fonctions de $C_b(\mathbb{R})$ définie par $h_k(x) = 1$ si $x \leq t$, $h_k(x) = 0$ si $x \geq t + 1/k$ et telle que h_k est affine entre t et $t + 1/k$. On a en particulier que $\mathbb{1}_{]-\infty, t]} \leq h_k \leq \mathbb{1}_{]-\infty, t+1/k]}$, et $(h_k)_{k \in \mathbb{N}_0}$ converge en tout point vers $\mathbb{1}_{]-\infty, t]}$. Ainsi, en utilisant l'hypothèse, il vient

$$F_{X_n}(t) = \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X_n)] \leq \mathbb{E}[h_k(X_n)] \longrightarrow \mathbb{E}[h_k(X)] \leq \mathbb{E}[\mathbb{1}_{]-\infty, t+1/k]}(X)] = F_X(t + 1/k)$$

lorsque n tend vers l'infini. Puisque t est un point de continuité de F_X , on obtient que

$$\limsup_{n \rightarrow +\infty} F_{X_n}(t) \leq \lim_{k \rightarrow +\infty} F_X(t + 1/k) = F_X(t).$$

On fait ensuite de même en considérant la suite $(h_k)_{k \in \mathbb{N}_0}$ de fonctions de $C_b(\mathbb{R})$ définie par $h_k(x) = 1$ si $x \leq t - 1/k$, $h_k(x) = 0$ si $x \geq t$ et telle que h_k est affine entre $t - 1/k$ et t , de sorte que

$$F_{X_n}(t) \geq \mathbb{E}[h_k(X_n)] \rightarrow \mathbb{E}[h_k(X)] \geq F_X(t - 1/k)$$

et on passe ensuite à la limite sur k .

4 \Rightarrow 2. Soit D l'ensemble des points de discontinuité de F_X . Par la Proposition 1.3.13, on sait que D est dénombrable. Par hypothèse, on sait que si $a, b \notin D$, alors

$$\mathbb{P}(X_n \in]a, b]) = F_{X_n}(b) - F_{X_n}(a) \longrightarrow F_X(b) - F_X(a) = \mathbb{P}(X \in]a, b]).$$

Soit $h \in C_c(\mathbb{R})$ et $\varepsilon > 0$. Comme la fonction h est uniformément continue sur \mathbb{R} , on peut trouver une fonction en escalier

$$g = \sum_{j=1}^N c_j \mathbb{1}_{]a_j, b_j]}$$

telle que $\sup_{x \in \mathbb{R}} |h(x) - g(x)| < \varepsilon$. De plus, comme D est dénombrable, on peut supposer que $a_j, b_j \notin D$ pour tout $j \in \{1, \dots, N\}$. Alors, on a

$$\mathbb{E}[g(X_n)] = \sum_{j=1}^N c_j \mathbb{P}(X_n \in]a_j, b_j]) \longrightarrow \sum_{j=1}^N c_j \mathbb{P}(X \in]a_j, b_j]) = \mathbb{E}[g(X)]$$

lorsque n tend vers l'infini et donc

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \\ & \leq \limsup_{n \rightarrow +\infty} (|\mathbb{E}[h(X_n)] - \mathbb{E}[g(X_n)]| + |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| + |\mathbb{E}[g(X)] - \mathbb{E}[h(X)]|) \\ & < 2\varepsilon, \end{aligned}$$

ce qui suffit puisque $\varepsilon > 0$ est arbitraire. ■

Remarque 5.5.5 L'équivalence entre la convergence en loi et la convergence des fonctions caractéristiques s'appelle le Théorème de Paul-Lévy.

Exemple 5.5.6 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans \mathbb{Z} . Alors, $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$ pour tout $k \in \mathbb{Z}$. Il suffit de considérer pour tout $k \in \mathbb{Z}$ une fonction continue h telle que $h(k) = 1$ et $h = 0$ en dehors de $[k - \varepsilon, k + \varepsilon]$.

Exemple 5.5.7 Si chaque variable aléatoire X_n admet une densité f_n et si la suite $(f_n)_{n \in \mathbb{N}}$ converge presque partout vers une fonction de densité f , alors $X_n \xrightarrow{\mathcal{L}} X$ où X est une variable aléatoire de densité f . En effet, si $h \in C_c(\mathbb{R})$, on a

$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq \int_{\mathbb{R}} |h(x)| |f_n(x) - f(x)| dx \leq \sup_{x \in \mathbb{R}} |h(x)| \int_{\mathbb{R}} |f_n(x) - f(x)| dx$$

et on a (en utilisant la relation $|a| = a + 2(-a)^+$)

$$\int_{\mathbb{R}} |f_n(x) - f(x)| dx = \int_{\mathbb{R}} f_n(x) - f(x) dx + 2 \int_{\mathbb{R}} (f(x) - f_n(x))^+ dx = 2 \int_{\mathbb{R}} (f(x) - f_n(x))^+ dx$$

puisque les densités ont une intégrale égale à 1. En utilisant l'inégalité $0 \leq (f - f_n)^+ \leq f$, le théorème de la convergence dominée donne

$$\int_{\mathbb{R}} (f(x) - f_n(x))^+ dx \longrightarrow 0$$

lorsque n tend vers l'infini, ce qui suffit.

Remarque 5.5.8 On vient de voir que la convergence des fonctions de densité implique la convergence en loi. Cette convergence est plus forte que la convergence en loi car la convergence en loi requiert la convergence de $\mathbb{P}(X_n \in]-\infty, b])$ vers $\mathbb{P}(X \in]-\infty, b])$ pour tout point $b \in \mathbb{R}^3$, alors que la convergence des densités donne une convergence "uniforme" c'est-à-dire telle que

$$\sup_{B \in \mathcal{B}} |\mathbb{P}(X_n \in B) - \mathbb{P}(X \in B)| \longrightarrow 0$$

puisque

$$\sup_{B \in \mathcal{B}} |\mathbb{P}(X_n \in B) - \mathbb{P}(X \in B)| = \sup_{B \in \mathcal{B}} \left| \int_B f_n(x) - f(x) dx \right| \leq \int_{\mathbb{R}} |f_n(x) - f(x)| dx \rightarrow 0.$$

Exemple 5.5.9 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires telles que X_n suit une loi uniforme sur $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$. Alors, $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une loi X uniforme sur $[0, 1]$. Cela découle de l'approximation de l'intégrale d'une fonction continue par les sommes de Riemann. En effet, si $h \in C_b(\mathbb{R})$, on a

$$\mathbb{E}[h(X_n)] = \sum_{j=1}^n h\left(\frac{j}{n}\right) \mathbb{P}\left(X_n = \frac{j}{n}\right) = \frac{1}{n} \sum_{j=1}^n h\left(\frac{j}{n}\right) \longrightarrow \int_0^1 h(x) dx = \mathbb{E}[h(X)].$$

Remarquons néanmoins que pour l'ensemble borélien $B = [0, 1] \cap \mathbb{Q}$, on a $\mathbb{P}(X_n \in B) = 1$ et $\mathbb{P}(X \in B) = 0$.

3. qui est nécessairement un point de continuité de F_X puisque la variable aléatoire est continue.

Exemple 5.5.10 Si $X_n \sim \mathcal{N}(0, \sigma_n^2)$ avec $\sigma_n \rightarrow 0$, alors $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire constante égale à 0. A nouveau, pour $B = \{0\}$, on a $\mathbb{P}(X_n \in B) = 0$ et $\mathbb{P}(X \in B) = 1$.

Remarque 5.5.11 (Théorème Portemanteau) Au vu des exemples précédents, de manière générale, si $X_n \xrightarrow{\mathcal{L}} X$, on ne peut pas affirmer que $\mathbb{P}(X_n \in B)$ converge vers $\mathbb{P}(X \in B)$ pour tout ensemble borélien B . On a cependant les équivalences suivantes (admis) :

- $X_n \xrightarrow{\mathcal{L}} X$,
- pour tout ouvert G de \mathbb{R} , $\liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$,
- pour tout fermé F de \mathbb{R} , $\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$,
- pour tout borélien A de \mathbb{R} tel que $\mathbb{P}(X \in A^\bullet) = 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$.

La définition de la convergence en loi nous donne gratuitement le résultat de continuité suivant.

Corollaire 5.5.12 Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et X une variable aléatoire. Si $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction continue et si $X_n \xrightarrow{\mathcal{L}} X$, alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

Démonstration : Il suffit de montrer que $\mathbb{E}[h(g(X_n))] \rightarrow \mathbb{E}[h(g(X))]$ pour tout $h \in C_b(\mathbb{R})$. Or, si $h \in C_b(\mathbb{R})$, on a également $h \circ g \in C_b(\mathbb{R})$ et la conclusion suit de l'hypothèse $X_n \xrightarrow{\mathcal{L}} X$. ■

Remarque 5.5.13 En général, on peut travailler uniquement avec des fonctions mesurables g telles que $\mathbb{P}(X \in D_g) = 0$ si D_g est l'ensemble des points de discontinuité de g .

Remarque 5.5.14 Ce corollaire, couplé à ce que nous savons déjà pour les autres modes de convergence, donne le résultat suivant. Si g est une fonction continue, alors

- $X_n \xrightarrow{\text{p.s.}} X \implies g(X_n) \xrightarrow{\text{p.s.}} g(X)$,
- $X_n \xrightarrow{\mathbb{P}} X \implies g(X_n) \xrightarrow{\mathbb{P}} g(X)$,
- $X_n \xrightarrow{\mathcal{L}} X \implies g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

En d'autres termes, les fonctions continues préservent les limites de suites de variables aléatoires en convergence presque sûre, en probabilité et en loi. Remarquons que la même chose n'est pas valide sans hypothèse supplémentaire pour la convergence en norme L^p car rien ne garantit que les variables aléatoires $g(X_n)$, $n \in \mathbb{N}$ et $g(X)$ soient dans L^p . Il faut donc supposer par exemple que g est bornée ou encore que g est Lipschitz.

Intéressons-nous à présent aux liens qu'il existe entre la convergence en loi et les autres types de convergence. Pour cela, nous devons imposer que les variables aléatoires $(X_n)_{n \in \mathbb{N}}$ soient définies sur le même espace probabilisé.

Proposition 5.5.15 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration : Fixons $h \in C_c(\mathbb{R})$ et $\varepsilon > 0$. Comme la fonction h est uniformément continue, il existe $\delta > 0$ tel que

$$|h(x) - h(y)| \leq \varepsilon \quad \text{si} \quad |x - y| \leq \delta.$$

Puisque $X_n \xrightarrow{\mathbb{P}} X$, on peut trouver un entier $N \in \mathbb{N}$ tel que

$$\mathbb{P}(|X_n - X| > \delta) \leq \frac{\varepsilon}{2B} \quad \forall n \geq N$$

avec $B = \sup |h|$. Pour tout $n \geq N$, on a alors

$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| &\leq |\mathbb{E}[(h(X_n) - h(X))\mathbf{1}_{\{|X_n - X| \leq \delta\}}]| + |\mathbb{E}[(h(X_n) - h(X))\mathbf{1}_{\{|X_n - X| > \delta\}}]| \\ &\leq \varepsilon + 2 \sup_{x \in \mathbb{R}} |h(x)| \mathbb{P}(|X_n - X| > \delta) \\ &\leq \varepsilon(1 + 2 \sup_{x \in \mathbb{R}} |h(x)|), \end{aligned}$$

ce qui suffit. ■

En général, la convergence en loi n'implique pas la convergence en probabilité. Il existe un cas particulier où les deux types de convergence sont équivalents.

Proposition 5.5.16 Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé et $c \in \mathbb{R}$ une constante. Si $X_n \xrightarrow{\mathcal{L}} c$, alors $X_n \xrightarrow{\mathbb{P}} c$.

Démonstration : Soit $\varepsilon > 0$ et soit h une fonction continue sur \mathbb{R} telle que $h(c) = 1$, $0 \leq h \leq 1$ et $h = 0$ en dehors de $[c - \varepsilon, c + \varepsilon]$. Puisque $X_n \xrightarrow{\mathcal{L}} c$, on a

$$\mathbb{E}[h(X_n)] \longrightarrow \mathbb{E}[h(c)] = h(c) = 1$$

lorsque n tend vers l'infini. Puisque $h \leq \mathbf{1}_{[c - \varepsilon, c + \varepsilon]}$, on en tire que

$$0 \leq \mathbb{P}(|X_n - c| > \varepsilon) = 1 - \mathbb{P}(X_n \in [c - \varepsilon, c + \varepsilon]) = 1 - \mathbb{E}[\mathbf{1}_{[c - \varepsilon, c + \varepsilon]}(X_n)] \leq 1 - \mathbb{E}[h(X_n)] \longrightarrow 0$$

lorsque n tend vers l'infini, d'où la conclusion. ■

Revenons au problème initial. Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires indépendantes et de même loi que $X \in L^1$. Par la loi forte des grands nombres, nous savons que

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\text{p.s.}} \mathbb{E}[X]$$

et on cherche à connaître l'ordre de grandeur de la différence $\frac{1}{n} \sum_{j=1}^n X_j - \mathbb{E}[X]$ lorsque n est grand. Si $X \in L^2$, on sait que

$$\mathbb{E}[(\bar{X}^{(n)} - \mathbb{E}[X])^2] = \mathbb{E} \left[\left(\frac{1}{n} \sum_{j=1}^n X_j - \mathbb{E}[X] \right)^2 \right] = \frac{1}{n^2} \sum_{j=1}^n \text{Var}[X_j] = \frac{\text{Var}[X]}{n}.$$

On peut donc s'attendre en moyenne à ce que l'ordre de grandeur de $\bar{X}^{(n)} - \mathbb{E}[X]$ soit de $\frac{1}{\sqrt{n}}$, ou de manière équivalente à ce que l'ordre de grandeur de $\sum_{j=1}^n X_j - n\mathbb{E}[X]$ soit de \sqrt{n} . Le théorème central limite précise cela.

Théorème 5.5.17 (Théorème Central Limite) Soit $(X_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de loi commune $X \in L^2$ d'espérance μ et de variance σ^2 . Alors

$$\sqrt{n}(\bar{X}^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Démonstration : Comme d'habitude, quitte à remplacer X par $X - \mathbb{E}[X]$, on peut supposer que les variables aléatoires sont centrées. Notons φ la fonction caractéristique de X . Comme les variables aléatoires X_n sont i.i.d. de même loi que X , on a

$$\varphi^{\sqrt{n}\bar{X}^{(n)}}(t) = \mathbb{E} \left[\exp \left(it \frac{X_1 + \dots + X_n}{\sqrt{n}} \right) \right] = \left(\varphi \left(\frac{t}{\sqrt{n}} \right) \right)^n.$$

Puisque $X \in L^2$, on sait que la fonction φ est de classe C^2 , et de plus on a $\varphi'(0) = i\mu = 0$ et $\varphi''(0) = -\mathbb{E}[X^2] = -\sigma^2$. En développant φ à l'ordre 2, on a

$$\varphi(x) = 1 + \varphi'(0)x + \varphi''(0)\frac{x^2}{2} + o(x^2) = 1 - \sigma^2\frac{x^2}{2} + o(x^2)$$

lorsque $x \rightarrow 0$. En particulier, pour tout t fixé, on a

$$\left(\varphi \left(\frac{t}{\sqrt{n}} \right) \right)^n = \left(1 - \sigma^2\frac{t^2}{2n} + \xi_n \right)^n = \exp \left(n \log \left(1 - \sigma^2\frac{t^2}{2n} + \xi_n \right) \right)$$

où $\xi_n = o(\frac{1}{n})$ ⁴ Par conséquent, on obtient

$$\lim_{n \rightarrow +\infty} \varphi^{\sqrt{n}\bar{X}^{(n)}}(t) = \lim_{n \rightarrow +\infty} \left(1 - \sigma^2\frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n = \exp \left(-\sigma^2\frac{t^2}{2} \right)$$

qui est la fonction caractéristique d'une loi $\mathcal{N}(0, \sigma^2)$. ■

Remarque 5.5.18 Montrons que l'on peut éviter le recours au logarithme complexe dans la preuve du TCL. Puisque

$$z^n - w^n = (z - w)(z^{n-1} + z^{n-2}w + \dots + zw^{n-2} + w^{n-1}),$$

alors si $z, w \in \mathbb{C}$ sont tels que $|z| \leq 1$ et $|w| \leq 1$ on a

$$|z^n - w^n| \leq n|z - w|.$$

4. On notera que ξ_n peut être complexe. Néanmoins, on peut étendre le logarithme défini sur $]0, +\infty]$ en un logarithme complexe défini sur $\mathbb{C} \setminus]-\infty, 0]$. En particulier, pour tout complexe z tel que $|z| < 1$, on peut considérer la série de puissances suivante

$$\log(1 + z) = \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} z^k.$$

Cela permet d'obtenir que $\log(1 + z) \sim z$ lorsque $z \rightarrow 0$.

En particulier, pour $z = \varphi(\frac{t}{\sqrt{n}})$ et $w = \exp(-\frac{\sigma^2 t^2}{2n})$, on a

$$\begin{aligned} \left| \varphi^{\sqrt{n}\bar{X}^{(n)}}(t) - \exp\left(-\frac{\sigma^2 t^2}{2}\right) \right| &\leq n \left| \varphi\left(\frac{t}{\sqrt{n}}\right) - \exp\left(-\frac{\sigma^2 t^2}{2n}\right) \right| \\ &= n \left| 1 - \sigma^2 \frac{t^2}{2n} + \xi_n - \exp\left(-\frac{\sigma^2 t^2}{2n}\right) \right| \\ &= n \left| \xi_n - \sum_{j=2}^{+\infty} \frac{(-1)^j \sigma^{2j} t^{2j}}{(2n)^j j!} \right| \\ &= n o\left(\frac{1}{n}\right) \end{aligned}$$

qui tend donc vers 0 lorsque n tend vers l'infini.

Remarque 5.5.19 On peut utiliser un argument similaire à celui de la preuve du TCL pour démontrer la loi faible des grands nombres dans le cas L^1 . Pour cela, on développe la fonction caractéristique de $\bar{X}^{(n)}$ à l'ordre 1, ce qui nous donne

$$\varphi^{\bar{X}^{(n)}}(t) = \left(\varphi^X\left(\frac{t}{n}\right) \right)^n = \left(1 + i \frac{t}{n} \mathbb{E}[X] + o\left(\frac{1}{n}\right) \right)^n.$$

On démontre ensuite que $\varphi^{\bar{X}^{(n)}}(t)$ converge vers la fonction caractéristique de $\delta_{\mathbb{E}[X]}$ en t , c'est-à-dire vers $\exp(it\mathbb{E}[X])$. On a donc la convergence en loi vers la constante $\mathbb{E}[X]$, ce qui implique la convergence en probabilité.

Par conséquent, si n est grand, la loi de la variable aléatoire Z_n définie par

$$Z_n = \frac{\bar{X}^{(n)} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{X}^{(n)} - \mu)}{\sigma}$$

est bien approximée par la loi $\mathcal{N}(0, 1)$. On pourra donc estimer des probabilités pour Z_n (et donc pour $\bar{X}^{(n)}$) en faisant comme s'il s'agissait d'une loi normale centrée et réduite. Ce qui rend le TCL si important est qu'il est vérifié quelle que soit la distribution de la variable $X \in L^2$.

Remarque 5.5.20 Quand deux variables sont proches en loi on écrit également $X \stackrel{\mathcal{L}}{\approx} Y$ ou même $X \approx Y$.

Présentons une conséquence importante du TCL qui permet d'approcher les lois binomiales par des lois normales.

Corollaire 5.5.21 (Théorème de Moivre-Laplace) *Supposons que $(X_n)_{n \in \mathbb{N}_0}$ est une suite de variables aléatoires telle que $X_n \sim \text{Bin}(n, p)$. Alors, on a*

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

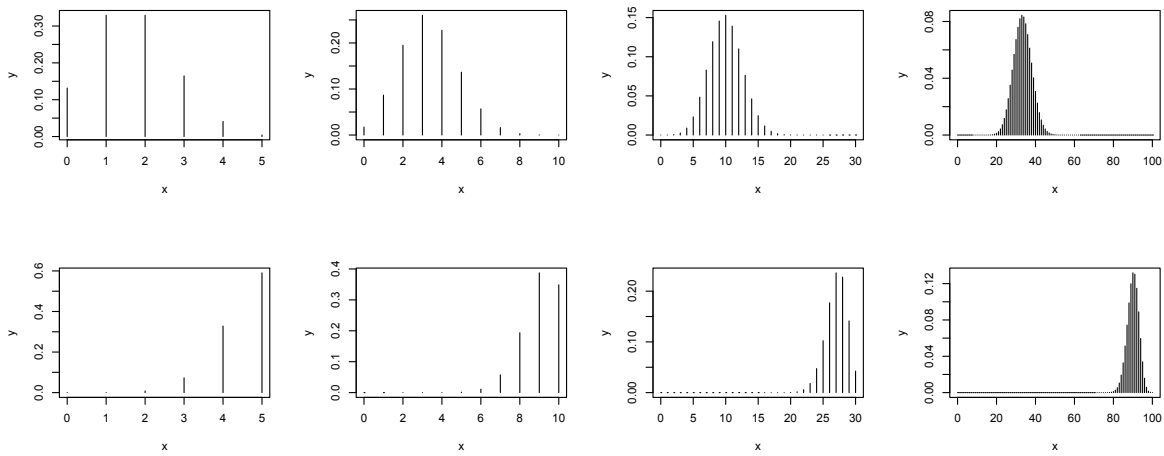


FIGURE 5.1 – Convergence vers la loi normale : diagrammes en bâtons des lois binomiales de paramètres (n, p) pour $n = 5, 10, 30, 100$ (de gauche à droite) et $p = \frac{1}{3}, \frac{9}{10}$ (haut et bas).

Démonstration : Soit $(Y_n)_{n \in \mathbb{N}_0}$ une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre p . Alors, on sait que $Y_1 + \dots + Y_n \sim \text{Bin}(n, p)$ et donc $\bar{Y}^{(n)}$ a la même loi que $\frac{X_n}{n}$. Le théorème Central Limite permet d’écrire

$$\frac{\bar{Y}^{(n)} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

et donc

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

■

Exemple 5.5.22 Si on considère le lancer d’une pièce de monnaie et l’événement “A= Face”, alors $p = \frac{1}{2}$. On souhaite calculer la probabilité d’obtenir entre 40 et 60 fois “Face” en 100 lancers d’une pièce de monnaie. Soit donc $Y \sim \text{Bin}(100, 1/2)$ la variable aléatoire qui compte le nombre de fois qu’on a obtenu “ Face”. On a $Y \approx \mathcal{N}(50, 25)$ et on calcule

$$\begin{aligned} \mathbb{P}[40 < Y \leq 60] &= \mathbb{P}\left(\frac{40 - 50}{5} < \frac{Y - 50}{5} \leq \frac{60 - 50}{5}\right) \\ &= \mathbb{P}\left(-2 < \frac{Y - 50}{5} \leq 2\right) \\ &\approx \mathbb{P}\left(-2 < Z \leq 2\right) \\ &\approx 0.9545 \end{aligned}$$

où $Z \sim \mathcal{N}(0, 1)$. Il y a donc à peu près 95.45% de chance qu’on ait entre 40 et 60 fois “Face” en

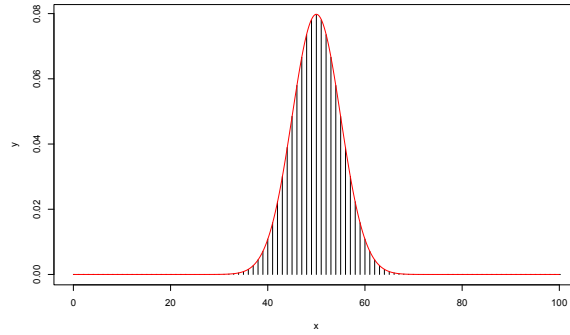


FIGURE 5.2 – Diagramme en bâtons de la loi $\text{Bin}(100, 1/2)$ et densité de la loi normale $\mathcal{N}(50, 25)$.

100 lancers d’une pièce de monnaie. On aurait pu calculer cette valeur de manière exacte :

$$\begin{aligned} \mathbb{P}(40 < Y \leq 60) &= \mathbb{P}(Y = 41) + \mathbb{P}(Y = 42) + \dots + \mathbb{P}(Y = 60) \\ &= \sum_{k=41}^{60} C_{100}^k \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{100-k} \approx 0.9540. \end{aligned}$$

Cette valeur exacte est très proche de l’approximation fournie par le TCL.

Le cas des variables aléatoires i.i.d. est extrêmement restreint et pour beaucoup d’applications on a besoin de quelque chose de beaucoup plus général. Présentons une condition qui permet de traiter le cas de variables aléatoires indépendantes non-nécessairement identiquement distribuées et considérons $S_n = X_{n,1} + \dots + X_{n,n}$ où $(X_{n,j})_{n \in \mathbb{N}_0, 1 \leq j \leq n}$ est une suite triangulaire, i.e. de la forme

$$\begin{array}{cccc} X_{1,1} & & & \\ X_{2,1} & X_{2,2} & & \\ X_{3,1} & X_{3,2} & X_{3,3} & \\ & \vdots & & \end{array}$$

Théorème 5.5.23 (Lindeberg-Feller) Soit $(X_{n,j})_{n \in \mathbb{N}_0, 1 \leq j \leq n}$ une suite triangulaire de variables aléatoires centrées et indépendantes de L^2 . Notons $\text{Var}[X_{n,j}] = \sigma_{n,j}^2$. Soit

$$Z_n = \sum_{j=1}^n X_{n,j}$$

et

$$B_n^2 = \text{Var}[Z_n] = \sum_{j=1}^n \sigma_{n,j}^2.$$

La condition Lindeberg-Feller (LF) est définie par

$$\forall \varepsilon > 0 \quad \frac{1}{B_n^2} \sum_{j=1}^n \mathbb{E} \left[X_{n,j}^2 \mathbf{1}_{\{|X_{n,j}| \geq \varepsilon B_n\}} \right] \rightarrow 0 \text{ lorsque } n \rightarrow +\infty. \quad (\text{LF})$$

On a

- si (LF) est vérifiée, alors $\frac{Z_n}{B_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.
- si $\max_{1 \leq j \leq n} \frac{\sigma_{n,j}^2}{B_n^2} \rightarrow 0$ quand $n \rightarrow +\infty$ et $\frac{Z_n}{B_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$, alors (LF) est vérifiée.

Exemple 5.5.24 Le TCL s'obtient en prenant $X_{n,j} = \frac{1}{\sqrt{n}}(X_j - \mu)$, auquel cas $\sigma_{n,j}^2 = \frac{\sigma^2}{n}$, $B_n^2 = \sigma^2$ et la condition (LF) est vérifiée puisque pour tout $\varepsilon > 0$

$$\frac{1}{\sigma^2} \sum_{j=1}^n \mathbb{E} \left[\frac{1}{n} (X_j - \mu)^2 \mathbf{1}_{\left\{ \frac{1}{\sqrt{n}} |X_j - \mu| \geq \varepsilon \sigma^2 \right\}} \right] = \frac{1}{\sigma^2} \mathbb{E} \left[(X - \mu)^2 \mathbf{1}_{\{|X - \mu| \geq \varepsilon \sigma^2 \sqrt{n}\}} \right]$$

qui tend bien vers 0 lorsque n tend vers l'infini par le théorème de la convergence monotone.

Exemple 5.5.25 *Inégalité de Le Cam.* Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes de loi Bernoulli de paramètres p_1, p_2, \dots, p_n respectivement. Soit $S_n = \sum_{j=1}^n X_j$, $\lambda = \sum_{j=1}^n p_j$ et $Z \sim \text{Poi}(\lambda)$. Dans cet exemple, nous allons montrer que

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(Z \in A)| \leq \sum_{j=1}^n p_j^2$$

pour tout $A \in \mathcal{B}$. Pour ce faire nous allons utiliser la méthode du *couplage*. Soient U_1, \dots, U_n des variables aléatoires indépendantes de loi uniforme sur $(0, 1)$ et définissons $X_j = \mathbf{1}_{\{U_j > 1 - p_j\}}$, $Y_j = 0$ si $U_j < e^{-p_j}$ et Y_j défini à partir de U_j de façon à suivre une loi de Poisson de paramètre p_j . Posons ensuite $Z = \sum_{j=1}^n Y_j$ et montrons (i) $|\mathbb{P}(S_n \in A) - \mathbb{P}(Z \in A)| \leq \mathbb{P}(S_n \neq Z)$; (ii) $\mathbb{P}(S_n \neq Z) \leq \sum_{j=1}^n \mathbb{P}(X_j \neq Y_j)$; (iii) $\mathbb{P}(X_j \neq Y_j) \leq p_j^2$. **Solution :** Soit $(U_i)_{i \geq 1}$ une suite de copies indépendantes d'une variable uniforme sur $[0, 1]$ et $X_i = \mathbb{I}(U_i > 1 - p_i)$. Alors X_1, \dots, X_n sont indépendantes de loi Bern(p_i). Soit maintenant F_i la fonction de répartition d'une loi de Poisson de paramètre p_i , et remarquons que $F_i(0), F_i(1), F_i(2), \dots$ découpe l'intervalle $(0, 1)$ en une infinité de morceaux disjoints. Prenons $Y_i = 0$ si $U_i < F_i(0)$, $Y_i = 1$ si $F_i(0) < U_i < F_i(1)$, $Y_i = 2$ si $F_i(1) < U_i < F_i(2)$ etc. Alors clairement $\mathbb{P}(Y_i = \ell) = F_i(\ell) - F_i(\ell - 1) = \mathbb{P}(Z_i = \ell)$ où $Z_i \sim \text{Poi}(p_i)$. Définissons, comme demandé, $Z = \sum_{i=1}^n Y_i$.

1. Pour montrer que $|\mathbb{P}(S_n \in A) - \mathbb{P}(Z \in A)| \leq \mathbb{P}(S_n \neq Z)$, supposons que $\mathbb{P}(S_n \in A) \geq \mathbb{P}(Z \in A)$. Alors

$$\begin{aligned} \mathbb{P}(S_n \in A) - \mathbb{P}(Z \in A) &\leq \mathbb{P}(S_n \in A) - \mathbb{P}(Z \in A \text{ et } S_n \in A) = \mathbb{P}(S_n \in A \text{ et } Z \notin A) \\ &\leq \mathbb{P}(S_n \neq Z) \end{aligned}$$

La conclusion suit.

2. Si $S_n \neq Z$ alors il doit y avoir au moins un i tel que $X_i \neq Y_i$. Par conséquent,

$$\mathbb{P}(S_n \neq Z) \leq \mathbb{P} \left(\bigcup_{i=1}^n (X_i \neq Y_i) \right) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i).$$

3. On a clairement

$$\mathbb{P}(X_i \neq Y_i) = 1 - \mathbb{P}(X_i = 0) - \mathbb{P}(Y_i = 1) = 1 - (1 - p_i) - p_i e^{-p_i} = p_i(1 - e^{-p_i}) \leq p_i^2.$$

En combinant ces dernières, on a

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(Z \in A)| \leq \mathbb{P}(S_n \neq Z) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) \leq \sum_{i=1}^n p_i^2.$$

5.6 Extension au cas multivarié et théorèmes de Slutsky

Les résultats présentés concernant la convergence en loi s'adaptent facilement dans le cas de vecteurs aléatoires. En particulier, le TCL prend la forme suivante.

Théorème 5.6.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires i.i.d. de vecteur moyen μ et de matrice de variance-covariance Σ . Alors

$$\sqrt{n}(\bar{X}^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma).$$

Nous ne faisons pas la preuve ici car elle est en tout point identique au cas de la dimension 1, en utilisant la formule de Taylor dans \mathbb{R}^d . Les résultats de la Proposition 5.5.4 restent valides et la preuve du théorème de Paul-Lévy (convergence des fonctions caractéristiques équivalente à la convergence en loi) dans le cas multivarié se démontre également de manière semblable, en remplaçant la variable aléatoire $Y \sim \mathcal{N}(0, \sigma^2)$ par un vecteur aléatoire $Y \sim \mathcal{N}(0, \sigma^2 \text{Id})$.

Lorsque l'on travaille avec une suite $(X_n)_{n \in \mathbb{N}}$ de vecteurs aléatoires, il est naturel de s'intéresser au lien qui existe entre la convergence en loi de la suite des vecteurs et la convergence en loi des suites de marginales. Remarquons que si $X_n \xrightarrow{\mathcal{L}} X$, alors puisque $p_j : x \in \mathbb{R}^d \mapsto x_j$ est continue et $p_j(X_n) = X_{n,j}$, on a la convergence en loi des marginales. La réciproque n'est pas vérifiée en général.

Exemple 5.6.2 Soit $U \sim \text{Unif}([0, 1])$. Considérons les suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ définies par $X_n = U$ pour tout $n \in \mathbb{N}$, et $Y_{2k} = U$, $Y_{2k+1} = 1 - U$, alors $X_n \xrightarrow{\mathcal{L}} U$ et $Y_n \xrightarrow{\mathcal{L}} U$ mais on montre facilement que le vecteur (X_n, Y_n) ne converge pas en loi.

Comme les conditions $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ n'impliquent pas que la suite de vecteurs aléatoires (X_n, Y_n) converge en loi vers (X, Y) , la convergence en loi des marginales n'est pas suffisante pour avoir la convergence en loi des vecteurs aléatoires. Le résultat suivant montre que pour avoir la convergence en loi d'une suite de vecteurs aléatoires, il faut avoir la convergence en loi de chacune des marginales mais aussi de toutes leurs combinaisons linéaires.

Corollaire 5.6.3 (Cramer-Wold) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires de \mathbb{R}^d . Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $t^T X_n \xrightarrow{\mathcal{L}} t^T X$ pour tout $t \in \mathbb{R}^d$.

Démonstration : Si $X_n \xrightarrow{\mathcal{L}} X$, alors on utilise le théorème de continuité pour la fonction continue $h : (x_1, \dots, x_d) \in \mathbb{R}^d \mapsto \sum_{j=1}^d t_j x_j$. Réciproquement, on utilise la caractérisation de la convergence en loi via les fonctions caractéristiques puisque pour tout $t \in \mathbb{R}^d$ et tout $x \in \mathbb{R}$, on

sait que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\exp \left(ix \sum_{j=1}^d X_{n,j} t_j \right) \right] = \mathbb{E} \left[\exp \left(ix \sum_{j=1}^d X_j t_j \right) \right].$$

Il suffit alors de prendre $x = 1$ puisqu'on retrouve alors les fonctions caractéristiques de X_n et de X en t . ■

Nous avons déjà vu que la convergence en loi avait de mauvaises propriétés vis-à-vis de la somme ou du produit. Nous allons à présent montrer qu'on peut avoir de meilleures propriétés si une des deux suites converge vers une constante.

Théorème 5.6.4 (Slutsky) Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de vecteurs aléatoires sur \mathbb{R}^d et $\mathbb{R}^{d'}$ respectivement. Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où $c \in \mathbb{R}^{d'}$, alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$.

Démonstration : On va montrer que la suite des fonctions caractéristiques de (X_n, Y_n) converge vers la fonction caractéristique de (X, c) . Soit $(x, y) \in \mathbb{R}^d \times \mathbb{R}^{d'}$. On a

$$\begin{aligned} \left| \varphi^{(X_n, Y_n)}(x, y) - \varphi^{(X, c)}(x, y) \right| &\leq \left| \varphi^{(X_n, Y_n)}(x, y) - \varphi^{(X_n, c)}(x, y) \right| + \left| \varphi^{(X_n, c)}(x, y) - \varphi^{(X, c)}(x, y) \right| \\ &= \left| \mathbb{E} \left[e^{i(X_n^T x + Y_n^T y)} - e^{i(X_n^T x + c^T y)} \right] \right| + \left| \mathbb{E} \left[e^{i(X_n^T x + c^T y)} - e^{i(X^T x + c^T y)} \right] \right| \\ &\leq \mathbb{E} \left[\left| e^{iX_n^T x} (e^{iY_n^T y} - e^{ic^T y}) \right| \right] + \left| e^{ic^T y} \mathbb{E} \left[e^{iX_n^T x} - e^{iX^T x} \right] \right| \\ &\leq \mathbb{E} \left[\left| e^{iY_n^T y} - e^{ic^T y} \right| \right] + \left| \varphi^{X_n}(x) - \varphi^X(x) \right|. \end{aligned}$$

Par hypothèse, on sait que le deuxième terme $|\varphi^{X_n}(x) - \varphi^X(x)|$ de la majoration tend vers 0 lorsque n tend vers l'infini. Il suffit donc de montrer que le premier terme de la majoration tend également vers 0. Soit $\varepsilon > 0$. Par continuité de l'application $t \in \mathbb{R}^{d'} \mapsto e^{it^T y}$ en c , il existe $\delta > 0$ tel que

$$|e^{it^T y} - e^{ic^T y}| < \varepsilon \quad \text{si} \quad \|t - c\| < \delta.$$

Comme la convergence en loi vers une constante implique la convergence en probabilité, il existe $N \in \mathbb{N}$ tel que

$$\mathbb{P}(\|c^T - Y_n\| \geq \delta) \leq \varepsilon$$

si $n \geq N$. Par conséquent, si $n \geq N$, on a

$$\begin{aligned} \mathbb{E} \left[\left| e^{iY_n^T y} - e^{ic^T y} \right| \right] &= \mathbb{E} \left[\left| e^{iY_n^T y} - e^{ic^T y} \right| \mathbf{1}_{\{\|c^T - Y_n\| < \delta\}} \right] + \mathbb{E} \left[\left| e^{iY_n^T y} - e^{ic^T y} \right| \mathbf{1}_{\{\|c^T - Y_n\| \geq \delta\}} \right] \\ &\leq \varepsilon + 2\mathbb{P}(\|c^T - Y_n\| \geq \delta) \\ &\leq 3\varepsilon, \end{aligned}$$

ce qui permet de conclure. ■

Corollaire 5.6.5 Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de vecteurs aléatoires sur \mathbb{R}^d .

1. Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où $c \in \mathbb{R}^d$, alors $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$ et $X_n^T Y_n \xrightarrow{\mathcal{L}} X^T c$.
2. Si $X_n \xrightarrow{\mathcal{L}} X$ et si $Y_n - X_n \xrightarrow{\mathcal{L}} 0$, alors $Y_n \xrightarrow{\mathcal{L}} X$.

Démonstration : 1. En utilisant le théorème de Slutsky 5.6.4, on obtient que $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$. Comme les fonctions $(x, y) \mapsto x + y$ et $(x, y) \mapsto x^T y$ sont continues, le résultat découle du second point du Corollaire 5.5.12.

2. Posons $Z_n = Y_n - X_n$ pour tout $n \in \mathbb{N}$. Alors, on a par hypothèse $Z_n \xrightarrow{\mathcal{L}} 0$ et par le premier point, $Y_n = X_n + Z_n \xrightarrow{\mathcal{L}} X + 0 = X$. ■

Théorème 5.6.6 (Cramér – Delta méthode) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires telle que $\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{L}} X$, où $\mu \in \mathbb{R}$. Alors,

- $X_n \xrightarrow{\mathbb{P}} \mu$
- si $g : \mathbb{R} \rightarrow \mathbb{R}$ est une application continue de classe C^1 dans un voisinage de μ , on a

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{L}} g'(\mu)X.$$

En particulier, si $\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$, alors

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(g'(\mu))^2).$$

Démonstration : Pour le premier point, on considère la suite de variables aléatoires $Y_n = \frac{1}{\sqrt{n}}$. Trivialement, on a $Y_n \xrightarrow{\mathcal{L}} 0$. Le théorème de Slutsky 5.6.4 implique que $\sqrt{n}(X_n - \mu)Y_n \xrightarrow{\mathcal{L}} 0$, donc on a également la convergence en probabilité puisque 0 est constant. Pour conclure, on remarque que

$$\sqrt{n}(X_n - \mu)Y_n = (X_n - \mu).$$

Pour le second point, on utilise le développement de Taylor de g en μ pour écrire

$$g(x) = g(\mu) + (x - \mu)g'(\mu) + r(x)$$

où $r(x) = o(x - \mu)$ si $x \rightarrow \mu$. Pour tout $n \in \mathbb{N}$, considérons la variable aléatoire Y_n définie par

$$Y_n = \begin{cases} \frac{r(X_n)}{(X_n - \mu)} = \frac{g(X_n) - g(\mu)}{(X_n - \mu)} - g'(\mu) & \text{si } X_n \neq \mu \\ 0 & \text{si } X_n = \mu. \end{cases}$$

Montrons que $Y_n \xrightarrow{\mathbb{P}} 0$. Soit $\varepsilon > 0$. Alors il existe $\delta > 0$ tel que $|r(x)| < \varepsilon|x - \mu|$ si $|x - \mu| < \delta$. Ainsi,

$$\mathbb{P}(|Y_n| \geq \varepsilon) = \mathbb{P}(r(X_n) \geq \varepsilon|X_n - \mu|) \leq \mathbb{P}(|X_n - \mu| \geq \delta).$$

Or par le premier point, on sait que $X_n \xrightarrow{\mathbb{P}} \mu$ et donc le membre de droite converge vers 0 lorsque n tend vers l'infini. On en tire que $Y_n \xrightarrow{\mathbb{P}} 0$ et donc également $Y_n \xrightarrow{\mathcal{L}} 0$. En écrivant

$$\sqrt{n}(g(X_n) - g(\mu)) = \sqrt{n}(X_n - \mu)g'(\mu) + Y_n\sqrt{n}(X_n - \mu),$$

la conclusion suit par le Théorème de Slutsky. ■

Remarque 5.6.7 Le premier point nous redonne la loi faible des grands nombres dans le cas L^2 comme conséquence du TCL.

5.7 Convergence dans le cas gaussien

Dans cette courte section, nous étudions le cas de la convergence d'une suite de variables aléatoires gaussiennes. On montre en particulier que la convergence en loi est équivalente à la convergence des suites des espérances et des variances. Commençons par rappeler que si $(x_n)_{n \in \mathbb{N}}$ est une suite de \mathbb{R} , alors un réel x est une *valeur d'adhérence* de $(x_n)_{n \in \mathbb{N}}$ s'il existe une sous-suite de $(x_n)_{n \in \mathbb{N}}$ qui converge vers x .

Lemme 5.7.1 *Si $(x_n)_{n \in \mathbb{N}}$ est une suite bornée de \mathbb{R} qui possède une unique valeur d'adhérence x , alors $(x_n)_{n \in \mathbb{N}}$ converge vers x .*

Démonstration : Supposons que $(x_n)_{n \in \mathbb{N}}$ ne converge pas vers x . Alors il existe $\varepsilon > 0$ et une sous-suite $(n_k)_{k \in \mathbb{N}}$ tels que $|x_{n_k} - x| > \varepsilon$ pour tout k . La suite $(x_{n_k})_{k \in \mathbb{N}}$ est bornée et on peut donc en extraire une sous-suite qui converge. La limite de cette sous-suite ne peut pas être x , ce qui contredit l'hypothèse d'unicité de la valeur d'adhérence. ■

Proposition 5.7.2 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires telle que $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$. La suite $(X_n)_{n \in \mathbb{N}}$ converge en loi si et seulement si $\mu_n \rightarrow \mu \in \mathbb{R}$ et $\sigma_n^2 \rightarrow \sigma^2 \in [0, +\infty[$. La loi limite est alors une loi normale $\mathcal{N}(\mu, \sigma^2)$.*

Démonstration : Si $\mu_n \rightarrow \mu \in \mathbb{R}$ et $\sigma_n^2 \rightarrow \sigma^2 \in [0, +\infty[$, alors il est clair que

$$\lim_{n \rightarrow +\infty} \exp\left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$$

qui est la fonction caractéristique d'une loi normale $\mathcal{N}(\mu, \sigma^2)$. Ainsi, on a bien $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(\mu, \sigma^2)$. Réciproquement, si $X_n \xrightarrow{\mathcal{L}} X$, alors on a que

$$\lim_{n \rightarrow +\infty} \exp\left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right) = \varphi^X(t)$$

pour tout $t \in \mathbb{R}$ et on aura donc le résultat si on montre que $\mu_n \rightarrow \mu \in \mathbb{R}$ et $\sigma_n^2 \rightarrow \sigma^2 \in [0, +\infty[$. Remarquons que

$$|\varphi^X(t)| = \lim_{n \rightarrow +\infty} \left| \exp\left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right) \right| = \lim_{n \rightarrow +\infty} \exp\left(-\frac{1}{2}\sigma_n^2 t^2\right)$$

pour tout $t \in \mathbb{R}$. En évaluant en $t = 1$, on a

$$|\varphi^X(1)| = \lim_{n \rightarrow +\infty} \exp\left(-\frac{1}{2}\sigma_n^2\right)$$

et donc la suite $(\sigma_n)_{n \in \mathbb{N}}$ converge. De plus, sa limite est nécessairement finie car sinon, cela impliquerait que pour tout $t \neq 0$

$$\lim_{n \rightarrow +\infty} \exp\left(-\frac{1}{2}\sigma_n^2 t^2\right) = 0 = |\varphi^X(t)|,$$

ce qui contredit la continuité de φ^X en 0 puisque $\varphi^X(0) = 1$. Notons σ^2 la limite de la suite $(\sigma_n)_{n \in \mathbb{N}}$. Montrons à présent que la suite $(\mu_n)_{n \in \mathbb{N}}$ converge. Pour cela, on commence par montrer

qu'elle est bornée. En effet, si ce n'est pas le cas, il existe une sous-suite $m_k \rightarrow +\infty$ telle que $\mu_{m_k} \rightarrow \pm\infty$. Supposons que $\mu_{m_k} \rightarrow +\infty$, le raisonnement étant similaire pour $-\infty$. Pour tout point de continuité x de F_X , on a

$$\mathbb{P}(X > x) = 1 - F_X(x) = 1 - \lim_{k \rightarrow +\infty} F_{X_{n_k}}(x) = \lim_{k \rightarrow +\infty} \mathbb{P}(X_{n_k} > x) \geq \frac{1}{2}$$

puisque si k est suffisamment grand, $x < \mu_{n_k}$ (et que l'espérance d'une loi normale est sa médiane). En prenant x arbitrairement grand (ce qui est possible puisque F_X admet seulement un nombre dénombrable de points de non-continuité), on obtient

$$\mathbb{P}(X = +\infty) \geq \frac{1}{2},$$

ce qui est impossible puisque $\mathbb{P}(X \in \mathbb{R}) = \lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in \mathbb{R}) = 1$. La suite $(\mu_n)_{n \in \mathbb{N}}$ est donc bornée. Ainsi, elle possède une sous-suite qui converge. Montrons que $(\mu_n)_{n \in \mathbb{N}}$ ne peut pas posséder deux valeurs d'adhérence. En effet, si μ et ν sont deux valeurs d'adhérence, alors on doit avoir $\varphi^X(t) = \exp(i\mu t - \frac{1}{2}\sigma^2 t^2) = \exp(i\nu t - \frac{1}{2}\sigma^2 t^2)$ pour tout $t \in \mathbb{R}$, donc $\exp(i\mu t) = \exp(i\nu t)$ pour tout $t \in \mathbb{R}$. On en tire⁵ que $\mu = \nu$ et on conclut en utilisant le Lemme 5.7.1. ■

5. par exemple en dérivant et en évaluant la dérivée en 0.

Annexe A

Introduction et rappels de théorie de la mesure

A.1 Approche intuitive de la notion de probabilité

Considérons une expérience aléatoire, c'est-à-dire une expérience dont l'issue n'est pas connue avec certitude. Nous supposons que, même si le résultat de l'expérience n'est pas connu, l'ensemble des résultats possibles est connu. On le note Ω et on l'appelle l'univers.

Exemple A.1.1

- Si on lance une pièce de monnaie, $\Omega = \{\text{Pile}, \text{Face}\}$.
- Si on lance deux dés distinguables, $\Omega = \{(j, k) : j, k \in \{1, \dots, 6\}\}$.
- Si on lance deux dés indistinguables, $\Omega = \{\{j, k\} : j, k \in \{1, \dots, 6\}\}$.
- Si on mesure le temps d'attente en minutes à un arrêt de bus, $\Omega = [0, +\infty[$.

On peut penser à une expérience aléatoire comme l'acte de *choisir un $\omega \in \Omega$ au hasard*. Un événement lié à cette expérience est une proposition liée au résultat de cette expérience, qui est soit vraie soit fausse à l'issue de l'expérience. On peut le voir comme une question que l'on peut se poser par rapport à cette expérience. Tout événement peut s'identifier à l'ensemble des résultats $\omega \in \Omega$ pour lesquels il se réalise. Tout événement est donc un sous-ensemble particulier de Ω .

Exemple A.1.2

- Si on lance deux dés distinguables, on peut s'intéresser à l'événement "*La somme des deux dés est égale à 8*" qui correspond au sous-ensemble

$$\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}.$$

- Lorsqu'on attend le bus, on peut s'intéresser à l'événement "*Le bus arrive dans 10 minutes maximum*" qui correspond au sous-ensemble $[0, 10]$.

Définition A.1.3 (Approche intuitive) La *probabilité* d'un événement A , que l'on notera $\mathbb{P}(A)$, est un nombre compris entre 0 et 1 qui mesure la confiance que l'on a en le fait que l'événement A va se produire.

Exemple A.1.4 Pour donner la probabilité de l'événement "*La somme des deux dés est égale à 8*", il suffit de passer en revue tous les cas : il y a 5 résultats favorables parmi les 36 résultats possibles. Ainsi, intuitivement, on s'attend à trouver que la probabilité est $5/36$.

Dans l'exemple précédent, il était facile d'estimer la probabilité de l'événement donné car nous avons un modèle probabiliste (équiprobable) en tête. Si on s'intéresse à présent à l'événement "Le bus arrive dans 10 minutes maximum", on ne pourra estimer sa probabilité que sur base de données. Ainsi, une manière de définir la probabilité d'un événement est d'utiliser une méthode statistique. On suppose que l'expérience d'intérêt est répétée un très grand nombre de fois sous les mêmes conditions. Si on s'intéresse à l'événement A , on note N_n le nombre de fois que l'événement s'est réalisé lors des n premières expériences. On en déduit la fréquence de réalisation de A lors des n premières expériences, donnée par

$$f_n = \frac{N_n}{n}.$$

On définit la probabilité de A comme la limite de la fréquence f_n lorsque le nombre d'expériences n tend vers l'infini :

$$\mathbb{P}(A) = \lim_{n \rightarrow +\infty} f_n = \lim_{n \rightarrow +\infty} \frac{N_n}{n}.$$

On parle de l'approche *fréquentiste* des probabilités.

Exemple A.1.5 On lance un dé et on considère l'événement $A = \text{"Le résultat est 1"}$. Après 15 lancers, on a

Résultat du dé	1	2	3	4	5	6
Effectif	2	4	3	3	1	2
Fréquence	0,133	0,266	0,2	0,2	0,066	0,133

et donc $f_{15} = 0,133$, et après 400 lancers

Résultat du dé	1	2	3	4	5	6
Effectif	67	56	68	62	80	67
Fréquence	0,1675	0,14	0,17	0,155	0,2	0,1674

et donc $f_{400} = 0,1675$ qui se rapproche de la valeur intuitive $1/6 \approx 0,166$, mais ce n'est pas le cas pour tous les résultats !

Même si cette définition est très intuitive et sera justifiée par la suite, certaines questions apparaissent :

- Comment être sûr que la limite existe ?
- Est-ce que la suite des fréquences va converger rapidement ?
- Comment être sûr que cette limite restera la même si on recommence l'expérience ?

L'objectif de la théorie des probabilités développée par Kolmogorov en 1933 est de donner un cadre formel via une axiomatique rigoureuse et efficace qui permet d'écrire toutes les questions liées au hasard. Le point de départ est une construction abstraite de la notion de *probabilité* comme une *mesure*. Intéressons-nous tout d'abord à l'ensemble de tous les événements liés à une expérience. On le note classiquement \mathcal{F} et on a donc $\mathcal{F} \subseteq \mathcal{P}(\Omega)$.

Exemple A.1.6 On lance une pièce infiniment souvent. On a donc $\Omega = \{\text{Pile}, \text{Face}\}^{\mathbb{N}_0}$. Pour tout $j \in \mathbb{N}_0$, on considère l'événement

$$A_j = \{\omega = (\omega_n)_{n \in \mathbb{N}_0} : \omega_j = \text{Pile}\},$$

i.e. A_j correspond à l'événement “On obtient Pile au $j^{\text{ème}}$ lancer”. Alors, on peut s'intéresser à l'événement $B_n = \cup_{j=1}^n A_j$ qui correspond au fait qu'on obtient au moins un Pile dans les n premiers lancers, tandis que $C_n = \cap_{j=1}^n A_j$ correspond à l'événement qu'on obtient Pile à chacun des n premiers lancers. On peut également imaginer des événements limites en prenant $\cup_{j=1}^{\infty} A_j$ (on obtient Pile à un moment) ou $\cup_{j=n}^{\infty} A_j$ (on obtient Pile après le $n^{\text{ème}}$ lancer) ou $\cap_{j=n}^{\infty} A_j$ (tous les lancers après le $n^{\text{ème}}$ sont sur le côté Pile). On peut aussi considérer l'événement

$$\Omega \setminus A_j = \{\omega = (\omega_n)_{n \in \mathbb{N}_0} : \omega_j = \text{Face}\},$$

qui correspond à l'événement “On obtient Face au $j^{\text{ème}}$ lancer”.

Le fait que n'importe quelle opération ensembliste (union dénombrable, intersection dénombrable, passage au complémentaire) permette d'engendrer de nouveaux événements nous suggère qu'il est nécessaire d'imposer qu'un “bon” système d'événements \mathcal{F} soit fermé sous les opérations ensemblistes. Naturellement, on souhaite également imposer à la probabilité \mathbb{P} de satisfaire les conditions minimales suivantes :

- pour tout $A \in \mathcal{F}$, $\mathbb{P}(A) \in [0, 1]$
- $\mathbb{P}(\Omega) = 1$
- pour tous $A, B \in \mathcal{F}$ tels que $A \subset B$, on a $\mathbb{P}(A) \leq \mathbb{P}(B)$
- si $(A_n)_n$ est une suite d'événements de \mathcal{F} deux à deux disjoints, alors

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n).$$

Exemple A.1.7 On considère une expérience dont les résultats forment un ensemble fini non-vide $\Omega = \{\omega_1, \dots, \omega_N\}$. On suppose que toutes les issues ont la même chance de se produire : on parle d'un contexte *équiprobable*. Naturellement, l'ensemble des événements liés à cette expérience est donné par $\mathcal{F} = \mathcal{P}(\Omega)$. On souhaite construire \mathbb{P} de manière à ce que chaque résultat ait la même probabilité de se produire, c'est-à-dire

$$\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_N\}).$$

Si p est cette probabilité commune, on doit imposer $Np = 1$ et donc $p = \frac{1}{N}$. Si on considère un événement $A = \{\omega_{i_1}, \dots, \omega_{i_k}\}$ de cardinalité k , on va avoir $\mathbb{P}(A) = \frac{k}{N}$, ce qu'on écrit souvent sous la forme

$$\mathbb{P}(A) = \frac{\text{Nombre de cas favorables pour } A}{\text{Nombre de cas possibles}} = \frac{\#A}{\#\Omega}.$$

Notons déjà que l'on parle de *distribution discrète uniforme* sur Ω .

Exemple A.1.8 Considérons à présent l'expérience qui consiste à choisir au hasard un nombre réel entre 0 et 1. On souhaite se placer dans le contexte équiprobable de l'exemple précédent. Intuitivement, cela se traduit en imposant que

$$\mathbb{P}(I_{a,b}) = b - a$$

si $I_{a,b}$ représente l'événement “Le nombre tiré appartient à l'intervalle $[a, b]$ ”. Autrement dit, on cherche une “mesure” qui à tout intervalle associe sa longueur. Il s'agit de la mesure de Lebesgue λ . Nous savons que si on souhaite que cette mesure ait de “bonnes propriétés”, il faut restreindre son ensemble de définition à la σ -algèbre de Borel. Ainsi, dans ce contexte, on a $\mathcal{F} = \mathcal{B}([0, 1])$ et $\mathbb{P} = \lambda|_{[0,1]}$ et on parle de *distribution continue uniforme* sur $[0, 1]$.

Comme le montre l'exemple précédent, si nous souhaitons que la mesure de probabilité ait des propriétés raisonnables, nous ne pouvons en général pas la définir sur tous les sous-ensembles de l'ensemble des résultats Ω .

A.2 σ -algèbres et applications mesurables

L'objectif de cette section est de rappeler les résultats essentiels liés aux σ -algèbres et aux applications mesurables.

Définition A.2.1 Soit E un ensemble non-vide. Une collection \mathcal{A} de sous-ensembles de E est une σ -algèbre (ou *tribu*) si

- $E \in \mathcal{A}$,
- pour tout $A \in \mathcal{A}$, on a $A^c \in \mathcal{A}$,
- pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Dans ce cas, on dit que (E, \mathcal{A}) est un *espace mesurable*.

Il est immédiat de vérifier que toute intersection non vide de σ -algèbres est encore une σ -algèbre. Dès lors, on peut adopter la définition suivante.

Proposition A.2.2 Si $\mathcal{U} \subseteq \mathcal{P}(E)$, il existe une plus petite (au sens de l'inclusion) σ -algèbre sur E qui contient \mathcal{U} . On l'appelle la σ -algèbre engendrée par \mathcal{U} et on la note $\sigma(\mathcal{U})$; cette σ -algèbre est obtenue par intersection de toutes les σ -algèbres qui contiennent \mathcal{U} .

Définition A.2.3 La σ -algèbre de Borel sur \mathbb{R} est

$$\mathcal{B}(\mathbb{R}) = \mathcal{B} = \sigma(\{]-\infty, b] : b \in \mathbb{R}\}),$$

c'est-à-dire la σ -algèbre sur \mathbb{R} engendrée par la collection des demi-droites réelles. Les éléments de \mathcal{B} sont appelés les ensembles *boréliens*.

Remarque A.2.4 La σ -algèbre de Borel est également engendrée par les ensembles ouverts, les ensembles fermés, les intervalles ouverts, fermés, semi-ouverts, etc. Mais elle n'est pas engendrée par les singletons.

Définition A.2.5 Soit $d \geq 1$. La σ -algèbre de Borel $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}^d$ sur \mathbb{R}^d est la σ -algèbre engendrée par la collection des demi-espaces de la forme $\prod_{k=1}^d]-\infty, b_k]$ pour $b_k \in \mathbb{R}$.

Remarque A.2.6 On peut définir la σ -algèbre de Borel sur n'importe quel espace topologique comme étant la σ -algèbre engendrée par les ouverts de la topologie. En particulier on peut également définir $\mathcal{B}(\mathbb{R})$ et montrer que $\mathcal{B}(\mathbb{R}) = \{B \cup C : B \in \mathcal{B}, C \subseteq \{-\infty, +\infty\}\}$.

Sans mention explicite du contraire, les espaces \mathbb{R}^d et $\overline{\mathbb{R}}$ seront munis de leur σ -algèbre de Borel. Rappelons à présent la notion de mesurabilité pour les applications qui joue un rôle central dans la théorie de l'intégration.

Définition A.2.7 Soient (E, \mathcal{A}) et (E', \mathcal{A}') deux espaces mesurables et $f : E \rightarrow E'$. On dit que f est *mesurable* (par rapport à \mathcal{A} et \mathcal{A}') si $f^{-1}(A') \in \mathcal{A}$ pour tout $A' \in \mathcal{A}'$.

La mesurabilité permet également d'introduire la notion de σ -algèbre engendrée par une application.

Proposition A.2.8 Soient E un ensemble, (E', \mathcal{A}') un espace mesurable et $f : E \rightarrow E'$. Alors

$$\{f^{-1}(A') : A' \in \mathcal{A}'\}$$

est la plus petite σ -algèbre sur E qui rend f mesurable. On l'appelle la σ -algèbre engendrée par f et on la note $f^{-1}(\mathcal{A}')$ ou $\sigma(f)$.

Par conséquent, dire que $f : E \rightarrow E'$ est mesurable par rapport à \mathcal{A} et \mathcal{A}' revient à imposer $\sigma(f) \subseteq \mathcal{A}$.

Remarque A.2.9 Plus généralement, si \mathcal{G} est une famille de fonctions de E dans un espace mesurable (E', \mathcal{A}') , on appelle σ -algèbre engendrée par \mathcal{G} la plus petite σ -algèbre qui rend mesurable toute fonction de \mathcal{G} . C'est donc la σ -algèbre engendrée par les ensembles de la forme $f^{-1}(A')$ pour $A' \in \mathcal{A}'$ et $f \in \mathcal{G}$. On la note $\sigma(\mathcal{G})$.

Nous allons à présent présenter quelques propriétés des fonctions mesurables. Le résultat suivant montre que pour qu'une fonction soit mesurable, il suffit de vérifier sa propriété caractéristique sur une famille génératrice de la σ -algèbre d'arrivée.

Proposition A.2.10 Soient (E, \mathcal{A}) et (E', \mathcal{A}') deux espaces mesurables et $f : E \rightarrow E'$. Supposons que $\mathcal{U} \subseteq \mathcal{P}(E')$ est tel que $\sigma(\mathcal{U}) = \mathcal{A}'$. Alors f est mesurable si et seulement si $f^{-1}(A') \in \mathcal{A}$ pour tout $A' \in \mathcal{U}$.

Intéressons-nous à présent au cas particulier d'applications à valeurs dans \mathbb{R} . Si (E, \mathcal{A}) est un espace mesurable et si $f : E \rightarrow \mathbb{R}$, alors f est mesurable si et seulement si

$$f^{-1}(]-\infty, b]) = \{x \in E : f(x) \leq b\} \in \mathcal{A}$$

pour tout $b \in \mathbb{R}$.

Proposition A.2.11 Soit (E, \mathcal{A}) un espace mesurable. Si $f : E \rightarrow \overline{\mathbb{R}}$ et $g : E \rightarrow \overline{\mathbb{R}}$ sont des applications mesurables et si $\alpha \in \mathbb{R}$, alors

- les ensembles

$$\{x \in E : f(x) < g(x)\}, \{x \in E : f(x) \leq g(x)\} \text{ et } \{x \in E : f(x) = g(x)\}$$

appartiennent à \mathcal{A} ,

- αf , $f + g$, $f - g$ et fg sont mesurables,
- $\max(f, g)$ et $\min(f, g)$ sont mesurables,
- si $g(x) \neq 0$ pour tout $x \in E$, alors $f/g : E \rightarrow \mathbb{R}$ est mesurable.

Terminons cette section en rappelant les résultats concernant les limites ponctuelles de fonctions mesurables.

Proposition A.2.12 Soit (E, \mathcal{A}) un espace mesurable. Si $(f_n)_{n \in \mathbb{N}}$ est une suite d'applications mesurables de E dans \mathbb{R} , alors

- $\sup_{n \in \mathbb{N}} f_n$ et $\inf_{n \in \mathbb{N}} f_n$ sont mesurables,
- $\limsup_{n \in \mathbb{N}} f_n$ et $\liminf_{n \in \mathbb{N}} f_n$ sont mesurables.

En particulier, si $(f_n)_{n \in \mathbb{N}}$ converge ponctuellement vers f , alors f est mesurable.

A.3 Mesures

Définition A.3.1 Soit (E, \mathcal{A}) un espace mesurable. Une application

$$\mu : \mathcal{A} \rightarrow [0, +\infty]$$

est une *mesure* sur (E, \mathcal{A}) si

- $\mu(\emptyset) = 0$,
- μ est σ -additif : si $(A_n)_{n \in \mathbb{N}}$ est une suite d'ensembles de \mathcal{A} deux à deux disjoints, alors

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

On dit que le triplet (E, \mathcal{A}, μ) est un *espace mesuré*.

Définition A.3.2 Une mesure est dite *finie* si $\mu(E) < +\infty$. Une mesure est dite σ -*finie* s'il existe une collection dénombrable $(A_k)_{k \in \mathbb{N}}$ d'éléments de \mathcal{A} tels que $E = \bigcup_{k \in \mathbb{N}} A_k$ et $\mu(A_k) < \infty$ pour tout $k \in \mathbb{N}$.

Définition A.3.3 Si μ est une mesure telle que $\mu(E) = 1$, on parle de *mesure de probabilité*. Le triplet (E, \mathcal{A}, μ) est alors appelé un espace probabilisé, E est l'univers et les éléments de \mathcal{A} les événements. On le note typiquement $(\Omega, \mathcal{F}, \mathbb{P})$.

Remarque A.3.4 Si μ est une mesure finie, on peut la standardiser en définissant $\tilde{\mu} = \frac{\mu}{\mu(E)}$. La mesure $\tilde{\mu}$ est alors une mesure de probabilité.

Exemple A.3.5 (Mesure de Dirac) Soient (E, \mathcal{A}) un espace mesurable et $x \in E$. La mesure de Dirac en x est définie par

$$\delta_x : \mathcal{A} \rightarrow \{0, 1\} : A \mapsto \delta_x(A) = \mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

où $\mathbb{1}_A$ est l'indicatrice de l'ensemble A . C'est une mesure de probabilité.

Exemple A.3.6 (Mesure de comptage) Soit (E, \mathcal{A}) un espace mesurable. On définit la mesure

$$\# : A \in \mathcal{A} \mapsto \begin{cases} \#A & \text{si } A \text{ est un ensemble fini} \\ +\infty & \text{si } A \text{ est un ensemble infini.} \end{cases}$$

Si E est fini, la mesure normalisée

$$A \mapsto \frac{\#A}{\#E}$$

est une mesure de probabilité qui modélise un tirage équiprobable.

Exemple A.3.7 (Mesure de Lebesgue) Si $E = \mathbb{R}$ et $\mathcal{A} = \mathbb{B}$, la mesure de Lebesgue λ sur (\mathbb{R}, \mathbb{B}) est la mesure construite à partir de la condition $\lambda([a, b]) = b - a$ (et définie proprement au cours de théorie de la mesure); ce n'est pas une mesure de probabilité. La mesure de Lebesgue est σ -finie.

Rappelons à présent les propriétés élémentaires des mesures.

Proposition A.3.8 Soit (E, \mathcal{A}, μ) un espace mesuré. Considérons $A, B \in \mathcal{A}$ et $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{A} .

1. *Monotonie* : si $A \subseteq B$, alors $\mu(A) \leq \mu(B)$. Si de plus $A \subseteq B$ et $\mu(A) < +\infty$, alors on a $\mu(B \setminus A) = \mu(B) - \mu(A)$.

2. *Sous-additivité dénombrable* : on a

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

3. *Continuité à gauche* : si la suite $(A_n)_{n \in \mathbb{N}}$ est croissante, alors

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mu(A_n).$$

4. *Continuité à droite* : si la suite $(A_n)_{n \in \mathbb{N}}$ est décroissante et s'il existe $n_0 \in \mathbb{N}$ pour lequel $\mu(A_{n_0}) < +\infty$, alors

$$\mu\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mu(A_n).$$

La proposition suivante est un outil de base de la théorie des probabilités. Dans ce contexte, il se lit de la manière suivante : si la série des probabilités d'une suite d'événements converge, alors la probabilité qu'une infinité d'entre eux se réalise simultanément est nulle. Nous y reviendrons.

Proposition A.3.9 (Lemme de Borel-Cantelli) Soient (E, \mathcal{A}, μ) un espace mesuré et une suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} . Si

$$\sum_{n \in \mathbb{N}} \mu(A_n) < +\infty,$$

alors

$$\mu\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0$$

où $\limsup_{n \rightarrow +\infty} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$.

A.4 Classes de Dynkin et Lemme de la classe monotone

Les classes de Dynkin fournissent un outil efficace pour tester l'égalité de mesures.

Définition A.4.1 Soit E un ensemble. Une collection \mathcal{D} de parties de E est une *classe de Dynkin* si

- $E \in \mathcal{D}$,
- pour tous $A, B \in \mathcal{D}$ tels que $A \subseteq B$, on a $B \setminus A \in \mathcal{D}$,
- pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{D} , on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}$.

Comme dans le cas de σ -algèbres, il est aisé de vérifier que l'intersection de classes de Dynkin est encore une classe de Dynkin. On peut donc définir la plus petite classe de Dynkin qui contient une collection \mathcal{C} de sous-ensembles de E , définie par l'intersection de toutes les classes de Dynkin qui contiennent \mathcal{C} . On l'appelle la *classe de Dynkin engendrée par \mathcal{C}* et on la note $\lambda(\mathcal{C})$. Bien sûr, toute σ -algèbre est une classe de Dynkin. Réciproquement, toute classe de Dynkin stable par intersection finie est une σ -algèbre. On peut dès lors montrer le résultat suivant.

Théorème A.4.2 (Dynkin) Si $\mathcal{C} \subseteq \mathcal{P}(E)$ est une famille stable par intersection finie, alors on a $\lambda(\mathcal{C}) = \sigma(\mathcal{C})$.

Le lemme de la classe monotone, dont la preuve est basée sur le Théorème de Dynkin, permet de prouver l'unicité de mesures.

Proposition A.4.3 (Lemme de la classe monotone) Soit (E, \mathcal{A}) un espace mesurable et soit \mathcal{C} une collection d'ensembles de E stable par intersection finie telle que $\sigma(\mathcal{C}) = \mathcal{A}$. Si μ_1 et μ_2 sont deux mesures σ -finies sur \mathcal{A} qui sont égales sur \mathcal{C} , alors $\mu_1 = \mu_2$.

Exemple A.4.4 (Unicité de la mesure de Lebesgue) Il existe au plus une mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout intervalle $[a, b]$, on ait $\lambda([a, b]) = b - a$. En effet, si λ' est une seconde mesure ayant la même propriété, on applique le Lemme de la classe monotone en prenant pour \mathcal{C} la collection des intervalles fermés qui engendrent $\mathcal{B}(\mathbb{R})$.

Exemple A.4.5 (Fonction de répartition) Si μ est une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors μ est entièrement déterminée par les valeurs de $\mu(]-\infty, b])$ avec $b \in \mathbb{R}$.

Dans le cas des mesures finies, qui seront notre objet d'étude principal pour la suite, on a la caractérisation suivante des fonctions de répartition.

Théorème A.4.6 Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Pour tout $x \in \mathbb{R}$, soit

$$F_\mu(x) = \mu(]-\infty, x]).$$

La fonction F_μ est croissante, bornée, continue à droite et satisfait $\lim_{x \rightarrow -\infty} F_\mu(x) = 0$. Inversement, si $F : \mathbb{R} \rightarrow \mathbb{R}^+$ est une fonction croissante, bornée, continue à droite telle que $\lim_{x \rightarrow -\infty} F(x) = 0$, alors il existe une unique mesure finie μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $F = F_\mu$. C'est la mesure de Lebesgue-Stieltjes associée à F .

Lorsque $F = F_\mu$ on note souvent

$$\int f d\mu = \int f dF.$$

On appelle cette dernière l'*intégrale de Stieltjes* par rapport à F . On a en particulier

$$\int_{]a,b]} dF(x) = F(b) - F(a) \quad \text{et} \quad \int_{[a,b]} dF(x) = F(b) - F(a^-).$$

Enfin, rappelons la notion d'ensemble négligeable.

Définition A.4.7 Soient (E, \mathcal{A}, μ) un espace mesuré. Un sous-ensemble N de E est μ -négligeable s'il existe un ensemble $A \in \mathcal{A}$ tel que $N \subseteq A$ et $\mu(A) = 0$. Une propriété a lieu μ -presque partout (on écrit μ -p.p.) si l'ensemble des points de E pour lesquels cette propriété n'est pas vérifiée est μ -négligeable.

A.5 Intégration et théorèmes limites

Dans cette section, nous rappelons la notion d'intégrale d'une application par rapport à une mesure ainsi que les théorèmes classiques de convergence. On suppose qu'un espace mesuré (E, \mathcal{A}, μ) est fixé. Naturellement, si $A \in \mathcal{A}$, on souhaite que l'intégrale de l'indicatrice de A par rapport à μ soit égal à $\mu(A)$, i.e.

$$\int \mathbb{1}_A d\mu = \mu(A).$$

Exemple A.5.1 Si λ est la mesure de Lebesgue, alors l'intégrale de $\mathbb{1}_{[a,b]}$ est simplement la longueur de l'intervalle $[a, b]$.

Exemple A.5.2 Si on considère sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ la mesure de comptage $\#$, alors pour tout sous-ensemble A de \mathbb{N} , l'intégrale de $\mathbb{1}_A$ par rapport à $\#$ est simplement la cardinalité de A .

Puisqu'on attend de l'intégrale qu'elle soit linéaire, on arrive naturellement à la définition suivante.

Définition A.5.3 Soient (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow [0, +\infty[$ une application de la forme

$$f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$$

où les ensembles A_k sont des éléments deux à deux disjoints de \mathcal{A} . L'intégrale de f par rapport à μ est définie par

$$\int f d\mu = \sum_{k=1}^n a_k \mu(A_k).$$

Les fonctions considérées dans la définition précédente sont les *fonctions mesurables simples positives*. L'ensemble de ces fonctions est noté $\mathcal{S}^+(E, \mathcal{A})$. Cette définition est ensuite étendue aux applications à valeurs positives de la manière suivante.

Définition A.5.4 Soient (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow [0, +\infty]$ une application mesurable. L'intégrale de f par rapport à μ , notée $\int f d\mu$, est définie par

$$\int f d\mu = \sup \left\{ \int g d\mu : g \in \mathcal{S}^+(E, \mathcal{A}), g \leq f \right\}.$$

Rappelons que si $f : E \rightarrow [0, +\infty]$ est une application mesurable, alors il existe une suite croissante $(f_n)_{n \in \mathbb{N}}$ de $\mathcal{S}^+(E, \mathcal{A})$ telle que $\lim_{n \rightarrow +\infty} f_n = f$ dans E . Dans ce cas, on peut aussi calculer l'intégrale de f via la relation

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Cette relation est parfois prise comme définition de l'intégrale. Enfin, la dernière étape consiste à étendre la définition de l'intégrale au cas des applications mesurables $f : E \rightarrow \overline{\mathbb{R}}$ arbitraires. Rappelons que les parties positives et négatives de f sont définies par $f^+ = \max(f, 0)$ et par $f^- = -\min(f, 0)$ respectivement. Ces applications sont mesurables si et seulement si f l'est.

Définition A.5.5 Soient (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \overline{\mathbb{R}}$ une application mesurable. Si les intégrales $\int f^+ d\mu$ et $\int f^- d\mu$ sont finies, alors on dit que f est *intégrable*. Dans ce cas, l'*intégrale* de f par rapport à μ , notée $\int f d\mu$, est définie par

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu. \quad (\text{A.1})$$

On dit que l'intégrale *existe* si $\int f^+ d\mu$ est fini ou si $\int f^- d\mu$ est fini. Dans ce cas, on définit également l'intégrale de f via (A.1).

Exemple A.5.6 Si λ est la mesure de Lebesgue, alors l'intégrale par rapport à λ étend la notion d'intégrale de Riemann.

Exemple A.5.7 Une application mesurable $f : E \rightarrow \mathbb{R}$ est intégrable par rapport à la mesure de Dirac δ_{x_0} pour $x_0 \in E$, et

$$\int f d\delta_{x_0} = f(x_0).$$

De manière générale, si $\mu = \sum_{n=1}^N \alpha_n \delta_{x_n}$ est une mesure discrète sur E , alors

$$\int f d\mu = \sum_{n=1}^N \alpha_n f(x_n).$$

Les premières propriétés de l'intégrale sont rassemblées dans le résultat ci-dessous.

Proposition A.5.8 Soient (E, \mathcal{A}, μ) un espace mesuré, $f, g : E \rightarrow \mathbb{R}$ deux applications intégrables et $\alpha \in \mathbb{R}$. Alors

- $|f| < +\infty$ μ -presque partout,
- αf est intégrable et $\int \alpha f d\mu = \alpha \int f d\mu$,
- $f + g$ est intégrable et $\int f + g d\mu = \int f d\mu + \int g d\mu$,
- si $f \leq g$ sur X , alors $\int f d\mu \leq \int g d\mu$,
- si $\int |f| d\mu = 0$, alors $f = 0$ μ -presque partout.

De plus, une application mesurable f est intégrable si et seulement si $|f|$ est intégrable. Dans ce cas,

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Remarque A.5.9 Soient $f, g : E \rightarrow \overline{\mathbb{R}}$ deux applications mesurables qui sont égales μ -p.p. Si $\int f d\mu$ existe, alors $\int g d\mu$ existe et

$$\int f d\mu = \int g d\mu.$$

Le résultat rappelé ci-dessous est connu sous le nom d'inégalité de Markov dans le contexte des probabilités. Nous y reviendrons.

Proposition A.5.10 Soient (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow [0, +\infty]$ une application mesurable. Pour tout $t > 0$, on pose $A_t = \{x \in E : f(x) \geq t\}$. Alors

$$\mu(A_t) \leq \frac{1}{t} \int_{A_t} f d\mu \leq \frac{1}{t} \int f d\mu.$$

Démonstration : Puisque $0 \leq t\mathbb{1}_{A_t} \leq f\mathbb{1}_{A_t} \leq f$, on a

$$\int t\mathbb{1}_{A_t} d\mu \leq \int_{A_t} f d\mu \leq \int f d\mu.$$

La conclusion s'obtient en remarquant que $\int t\mathbb{1}_{A_t} d\mu = t\mu(A_t)$. ■

Terminons cette section avec un rappel des théorèmes classiques de convergence.

Théorème A.5.11 (Convergence monotone) Soient (E, \mathcal{A}, μ) un espace mesuré, $(f_n)_{n \in \mathbb{N}}$ une suite croissante d'applications mesurables de X dans $[0, +\infty]$ et $f : E \rightarrow [0, +\infty]$ une application mesurable. Si $f = \lim_{n \rightarrow +\infty} f_n$, alors

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Remarque A.5.12 Si les relations $f_n \leq f_{n+1}$ pour tout $n \in \mathbb{N}$ et $f = \lim_{n \rightarrow +\infty} f_n$ n'ont lieu que μ -presque partout, alors la conclusion du théorème précédent reste valide.

Corollaire A.5.13 Soient (E, \mathcal{A}, μ) un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite d'applications mesurables de X dans $[0, +\infty]$. Alors

$$\int \sum_{n \in \mathbb{N}} f_n d\mu = \sum_{n \in \mathbb{N}} \int f_n d\mu.$$

Remarque A.5.14 Considérons le cas particulier de la mesure de comptage $\mu = \sum_{k=0}^{+\infty} \delta_k$ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On sait que

$$\int f d\mu = \sum_{k=0}^{+\infty} f(k).$$

Le corollaire précédent permet de démontrer que pour toute suite double $(a_{n,k})_{n,k \in \mathbb{N}}$ de réels positifs, on a

$$\sum_{n=0}^{+\infty} \left(\sum_{k=0}^{+\infty} a_{n,k} \right) = \sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} a_{n,k} \right).$$

Il suffit en effet de considérer une suite de fonctions $(f_n)_{n \in \mathbb{N}}$ telles que $f_n(k) = a_{n,k}$ pour tous $n, k \in \mathbb{N}$.

Théorème A.5.15 (Lemme de Fatou) Soient (E, \mathcal{A}, μ) un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite d'applications mesurables de E dans $[0, +\infty]$. Alors

$$\int \liminf_{n \rightarrow +\infty} f_n d\mu \leq \liminf_{n \rightarrow +\infty} \int f_n d\mu.$$

Théorème A.5.16 (Convergence dominée, de Lebesgue) Soient (E, \mathcal{A}, μ) un espace mesuré, $(f_n)_{n \in \mathbb{N}}$ une suite d'applications mesurables de E dans $\overline{\mathbb{R}}$ et f une application mesurable de E dans \mathbb{R} telle que $f = \lim_{n \rightarrow +\infty} f_n$. S'il existe une application intégrable $g : E \rightarrow [0, +\infty]$ telle que $|f_n| \leq g$ pour tout $n \in \mathbb{N}$, alors les applications f_n , $n \in \mathbb{N}$, et f sont intégrables et

$$\lim_{n \rightarrow +\infty} \int |f_n - f| d\mu = 0.$$

En particulier,

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Remarque A.5.17 Si les relations $|f_n| \leq g$ pour tout $n \in \mathbb{N}$ et $f = \lim_{n \rightarrow +\infty} f_n$ n'ont lieu que μ -presque partout, alors le résultat reste valide.

A.6 Mesure image

Le résultat suivant permet de transporter une mesure d'un espace vers un autre via une application mesurable.

Proposition A.6.1 Soient (E, \mathcal{A}, μ) un espace mesuré et (E', \mathcal{A}') un espace mesurable. Si $f : E \rightarrow E'$ est une application mesurable, alors l'application

$$\mu_f : \mathcal{A}' \rightarrow [0, +\infty] : A' \mapsto \mu(f^{-1}(A'))$$

est une mesure sur (E', \mathcal{A}') . On l'appelle la mesure image de μ par f .

Ces mesures images sont largement utilisées en théorie des probabilités. En effet, dans ce contexte, une application mesurable définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs réelles est appelée une *variable aléatoire*, souvent dénotée par X . On dit que \mathbb{P}_X est la *loi* ou la *distribution* de X . On retrouve

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

pour tout ensemble borélien A de \mathbb{R} , ce que l'on note souvent $\mathbb{P}(X \in A)$. Nous y reviendrons.

Exemple A.6.2 Considérons l'espace qui modélise le lancer d'un dé à six faces. Dans ce cas, on a $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mathbb{P}(A) = \frac{1}{6} \#A$ pour tout $A \subseteq \Omega$. Dans notre expérience, on ne s'intéresse pas au résultat du dé mais plutôt au fait que le résultat soit pair ou impair. Pour cela, on considère l'application $X : \Omega \rightarrow \{0, 1\}$ définie par $X(\omega) = 1$ si ω est pair, et $X(\omega) = 0$ si ω est impair. On vérifie immédiatement que

$$\mathbb{P}_X(\{0\}) = \mathbb{P}_X(\{1\}) = \frac{1}{2},$$

ce qui signifie qu'on a une chance sur deux d'obtenir un chiffre pair en jouant au dé.

L'intégrale par rapport à une mesure image s'effectue grâce au résultat suivant.

Théorème A.6.3 (de transfert) Soient (E, \mathcal{A}, μ) un espace mesuré, (E', \mathcal{A}') un espace mesurable et $f : E \rightarrow E'$ une application mesurable. Si $g : E' \rightarrow [0, +\infty]$ une application mesurable, alors

$$\int_{E'} g d\mu_f = \int_E g \circ f d\mu.$$

Si g est à valeurs réelles, alors g est μ_f -intégrable si et seulement si $g \circ f$ est μ -intégrable et dans ce cas, l'égalité précédente est satisfaite.

Démonstration : La preuve est classique. On commence par montrer que le résultat est valide pour $g = \mathbb{1}_B$ avec $B \in \mathcal{A}'$. On a

$$\int_{E'} g d\mu_f = \mu_f(B) = \mu(f^{-1}(B)) = \mu(\{x \in E : f(x) \in B\}) = \int_E \mathbb{1}_B \circ f d\mu$$

puisque $\mathbb{1}_B \circ f = 1$ si et seulement si $f(x) \in B$. Par linéarité de l'intégrale, on obtient le résultat pour toute fonction de $\mathcal{S}^+(E', \mathcal{A}')$. La théorème de la convergence monotone permet de montrer que l'égalité reste valide pour toute fonction mesurable à valeurs positives. Enfin, on obtient le résultat général en travaillant sur les parties positive et négative. ■

A.7 Le théorème de Radon-Nikodym

Etant donnée une mesure sur un espace, nous avons vu que l'on peut construire de nouvelles mesures en considérant les mesures images. Nous considérons à présent une deuxième manière de construire des mesures.

Proposition A.7.1 Soient (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow [0, +\infty]$ une application mesurable. On définit l'application

$$\nu : \mathcal{A} \rightarrow [0, +\infty] : A \mapsto \int_A f d\mu.$$

Alors ν est une mesure sur (E, \mathcal{A}) et on dit que f est une densité de ν par rapport à μ . On note $\nu = f \cdot \mu$.

Exemple A.7.2 En théorie des probabilités, les lois de variables aléatoires \mathbb{P}_X sont souvent définies comme des densités par rapport à la mesure de Lebesgue. Par exemple, pour la loi normale, on a

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Proposition A.7.3 Soient (E, \mathcal{A}, μ) un espace mesuré, g une application mesurable de E dans $[0, +\infty]$ et f une application mesurable de E dans \mathbb{R} . Alors f est $g \cdot \mu$ -intégrable si et seulement si fg est μ -intégrable et dans ce cas,

$$\int f d(g \cdot \mu) = \int fg d\mu.$$

Exemple A.7.4 Si X est une variable aléatoire sur un espace probablisé $(\Omega, \mathcal{A}, \mathbb{P})$ de loi \mathbb{P}_X , on définit l'espérance de X par

$$\mathbb{E}[X] = \int \text{id} d\mathbb{P}_X.$$

Si \mathbb{P}_X est une mesure qui possède une densité f par rapport à la mesure de Lebesgue, alors on retrouve

$$\mathbb{E}[X] = \int \text{id} d\mathbb{P}_X = \int \text{id} d(f \cdot \lambda) = \int \text{id} f d\lambda = \int_{\mathbb{R}} x f(x) dx.$$

Remarquons que la mesure $\nu = f \cdot \mu$ définie par

$$\nu : \mathcal{A} \rightarrow [0, +\infty] : A \mapsto \int_A f d\mu$$

satisfait la propriété suivante : si $A \in \mathcal{A}$ est tel que $\mu(A) = 0$, alors $f\mathbb{1}_A = 0$ μ -presque partout, d'où $\nu(A) = 0$.

Définition A.7.5 Soit (E, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures sur (E, \mathcal{A}) . On dit que ν est *absolument continue* par rapport à μ si pour tout $A \in \mathcal{A}$ tel que $\mu(A) = 0$, on a $\nu(A) = 0$. Dans ce cas, on écrit

$$\nu \ll \mu.$$

Une mesure sur $(\mathbb{R}^d, \mathcal{B}^d)$ est *absolument continue* si elle est absolument continue par rapport à la mesure de Lebesgue.

La Proposition A.7.1 fournit donc toute une classe de mesures absolument continues par rapport à μ , à savoir les mesures à densité. Le Théorème de Radon-Nikodym donne la réciproque de ce résultat.

Théorème A.7.6 (Théorème de Radon-Nikodym) Soit (E, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures σ -finies sur (E, \mathcal{A}) . Alors $\nu \ll \mu$ si et seulement si il existe une application mesurable $f : E \rightarrow [0, +\infty]$ telle que

$$\nu(A) = \int_A f d\mu$$

pour tout $A \in \mathcal{A}$, c'est-à-dire $\nu = f \cdot \mu$. De plus, si g est une autre application telle que $\nu = g \cdot \mu$, alors $f = g$ μ -presque partout.

La fonction apparaissant dans le Théorème de Radon-Nikodym (définie μ -presque partout) est notée $\frac{d\nu}{d\mu}$. Ainsi, avec les notations précédemment introduites, on a

$$\nu = \frac{d\nu}{d\mu} \cdot \mu.$$

On peut affiner le Théorème de Radon-Nikodym en donnant une décomposition de toute mesure σ -finie, appelée la décomposition de Lebesgue. Afin de présenter ce résultat, nous allons nous intéresser à la notion suivante.

Définition A.7.7 Soit (E, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures sur (E, \mathcal{A}) . On dit que μ et ν sont *mutuellement singulières* s'il existe un ensemble $A \in \mathcal{A}$ tel que

$$\mu(A) = 0 \quad \text{et} \quad \nu(A^c) = 0.$$

Dans ce cas, on écrit

$$\mu \perp \nu.$$

Une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est *singulière* si elle est mutuellement singulière avec la mesure de Lebesgue.

Exemple A.7.8 Si μ est une mesure discrète sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, c'est-à-dire $\mu = \sum_{n=1}^N \alpha_n \delta_{x_n}$, alors μ est singulière puisque $\lambda(D) = 0$ si $D = \{x_n : n \in \{1, \dots, N\}\}$.

Remarquons que dans ce cas, si $B \in \mathcal{A}$ est tel que $\nu(B) > 0$, alors $B \cap A \neq \emptyset$. Ainsi, ν concentre son effet sur l'ensemble A qui est de μ mesure nulle. À l'opposé, si $\nu \ll \mu$, alors ν n'a aucun effet sur les ensembles de μ mesure nulle. Il s'agit donc de deux notions qui s'excluent.

Théorème A.7.9 (Décomposition de Lebesgue) Soit (E, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures σ -finies sur (E, \mathcal{A}) . Il existe un unique couple (ν_a, ν_s) de mesures sur (E, \mathcal{A}) telles que

$$\nu_a \ll \mu, \quad \nu_s \perp \mu \quad \text{et} \quad \nu = \nu_a + \nu_s.$$

On appelle le couple (ν_a, ν_s) la décomposition de Lebesgue de ν par rapport à μ . On dit que ν_a est la partie absolument continue de ν par rapport à μ et que ν_s est la partie singulière de ν par rapport à μ .

Dans le cas où $\mu = \lambda$, on sait que toute mesure discrète sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est singulière. Néanmoins, toute mesure singulière n'est pas nécessairement discrète. La mesure ν_s apparaissant dans la décomposition de Lebesgue n'est pas nécessairement discrète. La décomposition la plus fine est proposée dans la Proposition C.3.11.

A.8 Mesure produit et théorème de Tonelli-Fubini

Soient (E, \mathcal{A}, μ) et (E', \mathcal{A}', ν) deux espaces mesurés. On souhaite munir le produit cartésien $E \times E'$ d'une mesure $\mu \times \nu$ telle que

$$\mu \times \nu(A \times A') = \mu(A)\nu(A')$$

pour tous $A \in \mathcal{A}, A' \in \mathcal{A}'$.

Définition A.8.1 Soient (E, \mathcal{A}) et (E', \mathcal{A}') deux espaces mesurables. Tout ensemble de la forme $A \times A'$ avec $A \in \mathcal{A}, A' \in \mathcal{A}'$ est appelé un *rectangle mesurable*. La σ -algèbre produit sur $E \times E'$ est la σ -algèbre engendrée par les rectangles mesurables. On la note $\mathcal{A} \otimes \mathcal{A}'$. C'est la plus petite σ -algèbre sur $E \times E'$ rendant les projections

$$\pi_1 : E \times E' \rightarrow E : (x, y) \mapsto x \quad \text{et} \quad \pi_2 : E \times E' \rightarrow E' : (x, y) \mapsto y$$

mesurables.

Mentionnons que cette dernière propriété sert de base pour la définition d'un produit infini de σ -algèbres.

Remarque A.8.2 On a $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

Etant donné un ensemble $H \subseteq E \times E'$, on s'intéresse aux sections définies par

$$H_x = \{y \in E' : (x, y) \in H\}$$

et

$$H^y = \{x \in E : (x, y) \in H\}$$

pour tous $x \in E, y \in E'$. Pour définir la mesure d'un ensemble de $E \times E'$, on souhaiterait prendre

$$\int_E \nu(H_x) d\mu(x) \quad \text{ou} \quad \int_{E'} \mu(H^y) d\nu(y).$$

Dans le cas particulier d'un rectangle mesurable $A \times A'$, puisqu'on a $(A \times A')_x = A' \mathbb{1}_A(x)$ et $(A \times A')^y = A \mathbb{1}_{A'}(y)$, on retrouverait bien

$$\int_E \nu((A \times A')_x) d\mu(x) = \int_E \nu(A') \mathbb{1}_A(x) d\mu(x) = \mu(A) \nu(A')$$

et

$$\int_{E'} \mu((A \times A')^y) d\nu(y) = \int_{E'} \mu(A) \mathbb{1}_{A'}(y) d\nu(y) = \mu(A) \nu(A').$$

On peut montrer que cette construction a du sens, c'est-à-dire que les ensembles H_x et H^y appartiennent à \mathcal{A}' et \mathcal{A} respectivement, que les applications $x \mapsto \nu(H_x)$ et $y \mapsto \mu(H^y)$ sont mesurables et enfin que les intégrales donnent bien des mesures qui coïncident.

Théorème A.8.3 Soient (E, \mathcal{A}, μ) et (E', \mathcal{A}', ν) deux espaces mesurés σ -finis. Il existe une unique mesure $\mu \times \nu$ sur l'espace mesurable $(E \times E', \mathcal{A} \otimes \mathcal{A}')$ telle que

$$\mu \times \nu(A \times A') = \mu(A) \nu(A')$$

pour tous $A \in \mathcal{A}, A' \in \mathcal{A}'$. Pour tout $H \in \mathcal{A} \otimes \mathcal{A}'$, on a

$$\mu \times \nu(H) = \int_E \nu(H_x) d\mu(x) = \int_{E'} \mu(H^y) d\nu(y).$$

On dit que $\mu \times \nu$ est le produit des mesures μ et ν sur $\mathcal{A} \otimes \mathcal{A}'$.

Le Théorème de Fubini permet de calculer des intégrales par rapport à une mesure produit grâce à des intégrales itérées.

Théorème A.8.4 (Fubini) Soient (E, \mathcal{A}, μ) et (E', \mathcal{A}', ν) deux espaces mesurés σ -finis et soit $f : E \times E' \rightarrow \overline{\mathbb{R}}$ une application $\mu \times \nu$ -intégrable.

1. L'application $z \in E' \mapsto f(x, z)$ est ν -intégrable pour μ -presque tout $x \in E$ et l'application $z \in E' \mapsto f(z, y)$ est μ -intégrable pour ν -presque tout $y \in E'$.
2. Les applications $x \in E \mapsto \int_{E'} f(x, y) d\nu(y)$ et $y \in E' \mapsto \int_E f(x, y) d\mu(x)$, bien définies sauf sur un ensemble mesurable de mesure nulle, sont μ -intégrable et ν -intégrable respectivement.

3. On a

$$\int_{E \times E'} f d\mu \times \nu = \int_E \left(\int_{E'} f(x, y) d\nu(y) \right) d\mu(x) = \int_{E'} \left(\int_E f(x, y) d\mu(x) \right) d\nu(y).$$

A.9 Espaces L^p

Dans toute cette section, on suppose qu'un espace mesuré (E, \mathcal{A}, μ) est fixé. On suppose de plus que la mesure μ est σ -finie. Cette hypothèse permet de simplifier la définition de l'espace $\mathcal{L}^\infty(E, \mathcal{A}, \mu)$; les résultats présentés restent valides pour une mesure quelconque.

Définition A.9.1 Si $p \geq 1$, on définit l'espace $\mathcal{L}^p(E, \mathcal{A}, \mu)$ comme l'ensemble des applications mesurables $f : E \rightarrow \mathbb{R}$ telles que $|f|^p$ est intégrable. L'espace $\mathcal{L}^\infty(E, \mathcal{A}, \mu)$ est défini comme l'ensemble des applications mesurables $f : E \rightarrow \mathbb{R}$ *essentiellement bornées*, c'est-à-dire telles qu'il existe $C > 0$ tel que

$$\{x \in E : |f(x)| > C\}$$

est μ -négligeable.

Cette définition s'étend naturellement au cas des applications à valeurs complexes. Ces espaces sont des espaces vectoriels. Pour chaque p , on définit une relation d'équivalence sur $\mathcal{L}^p(E, \mathcal{A}, \mu)$ en posant

$$f \sim g \iff f = g \text{ } \mu\text{-pp.}$$

Cela conduit à définir l'espace quotient

$$L^p(E, \mathcal{A}, \mu) = \mathcal{L}^p(E, \mathcal{A}, \mu) / \sim.$$

Un élément de $L^p(E, \mathcal{A}, \mu)$ est donc une classe de fonctions de $\mathcal{L}^p(E, \mathcal{A}, \mu)$ qui sont égales μ -pp. Dans la suite, par abus de notation, on identifiera toujours un élément de $L^p(E, \mathcal{A}, \mu)$ à un de ses représentants.

Définition A.9.2 Si $p \in [1, +\infty[$, on pose

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{\frac{1}{p}}$$

pour tout $f \in L^p(E, \mathcal{A}, \mu)$. De plus, on définit pour tout $f \in L^\infty(E, \mathcal{A}, \mu)$,

$$\|f\|_\infty = \inf\{C \geq 0 : \{x \in E : |f(x)| > C\} \text{ est } \mu\text{-négligeable}\}.$$

Nous allons à présent montrer que les applications $\|\cdot\|_p$ définissent des normes. Commençons par introduire l'exposant conjugué de p .

Définition A.9.3 Soit $p \in]1, +\infty[$. Alors $\frac{1}{p} \in]0, 1[$ et il existe un nombre réel q tel que

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Les nombres p et q sont appelés des *exposants conjugués*. On étend cette définition en im-

posant que les nombres 1 et $+\infty$ sont des exposants conjugués.

Remarquons que si p et q sont des exposants conjugués, alors

$$p + q = pq$$

et si p et q sont finis,

$$p = q(p - 1) \quad \text{et} \quad q = p(q - 1).$$

Lemme A.9.4 Soit $p \in]1, +\infty[$ et soit q son exposant conjugué. Alors

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

pour tous $x, y \geq 0$.

Démonstration : Bien sûr, on peut supposer que x et y sont non-nuls. En posant $u = x^p$ et $v = y^q$, il suffit de prouver que

$$u^{1/p}v^{1/q} \leq \frac{u}{p} + \frac{v}{q}$$

pour tous $u, v > 0$. En posant $t = \frac{u}{v}$, il suffit de montrer que

$$t^{1/p} \leq \frac{t}{p} + \frac{1}{q}$$

pour tout $t > 0$. Il suffit alors de remarquer que la fonction

$$t > 0 \mapsto \frac{t}{p} + \frac{1}{q} - t^{1/p}$$

atteint son minimum en $t = 1$ et que ce minimum vaut 0. ■

Proposition A.9.5 (Inégalité de Hölder) Soit $p \in [1, +\infty[$ et soit q son exposant conjugué. Si $f \in L^p(E, \mathcal{A}, \mu)$ et $g \in L^q(E, \mathcal{A}, \mu)$, alors $fg \in L^1(E, \mathcal{A}, \mu)$ et

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

Démonstration : Le cas où $p = 1$ et $q = +\infty$ est immédiat. Supposons donc que $p, q \in]1, +\infty[$. Si $\|f\|_p = 0$ ou $\|g\|_q = 0$, alors $fg = 0$ presque partout et le résultat est démontré. Sinon, on pose $\tilde{f} = \frac{f}{\|f\|_p}$ et $\tilde{g} = \frac{g}{\|g\|_q}$. Par le Lemme A.9.4, on a

$$|\tilde{f}(x)\tilde{g}(x)| \leq \frac{|\tilde{f}(x)|^p}{p} + \frac{|\tilde{g}(x)|^q}{q}$$

pour tout $x \in X$. Par conséquent, $\tilde{f}\tilde{g} \in L^1(E, \mathcal{A}, \mu)$ et

$$\int |\tilde{f}\tilde{g}| d\mu \leq \frac{1}{p} \int |\tilde{f}|^p d\mu + \frac{1}{q} \int |\tilde{g}|^q d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

On en tire que $fg \in L^1(E, \mathcal{A}, \mu)$ et que

$$\frac{1}{\|f\|_p \|g\|_q} \int |fg| d\mu \leq 1. \quad \blacksquare$$

Proposition A.9.6 (Inégalité de Minkowski) Soit $p \in [1, +\infty]$. Si $f, g \in L^p(E, \mathcal{A}, \mu)$, alors

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Démonstration : Supposons d'abord que $p = +\infty$. Soit $N_1 = \{x \in X : |f(x)| > \|f\|_\infty\}$ et $N_2 = \{x \in E : |g(x)| > \|g\|_\infty\}$. Alors $N_1 \cup N_2$ est μ -négligeable et on a

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty$$

pour tout $x \notin N_1 \cup N_2$. On en tire que $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

Supposons à présent que $p = 1$. Puisque $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ pour tout $x \in E$, on a

$$\|f + g\|_1 = \int |f + g| d\mu \leq \int |f| d\mu + \int |g| d\mu = \|f\|_1 + \|g\|_1.$$

Supposons finalement que $p \in]1, +\infty[$. Si $\|f + g\|_p = 0$, le résultat est clair. Sinon, soit q l'exposant conjugué de p . Alors $p = q(p - 1)$ et par conséquent,

$$|f + g|^p = (|f + g|^{p-1})^q.$$

On en tire que $|f + g|^{p-1} \in L^q(E, \mathcal{A}, \mu)$. En utilisant l'inégalité de Hölder, on obtient

$$\begin{aligned} \|f + g\|_p^p &= \int |f + g|^p d\mu \leq \int |f + g|^{p-1} (|f| + |g|) d\mu \\ &= \int |f + g|^{p-1} |f| d\mu + \int |f + g|^{p-1} |g| d\mu \\ &\leq \| |f + g|^{p-1} \|_q \|f\|_p + \| |f + g|^{p-1} \|_q \|g\|_p \\ &= (\|f\|_p + \|g\|_p) \| |f + g|^{p-1} \|_q \\ &= (\|f\|_p + \|g\|_p) \left(\int |f + g|^p d\mu \right)^{1-1/p} \end{aligned}$$

puisque $q(p - 1) = p$ et $1/q = 1 - 1/p$. En divisant par $(\int |f + g|^p d\mu)^{1-1/p}$, cette inégalité se réécrit

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

■

Annexe B

Notions fondamentales de la théorie des probabilités

B.1 Espaces probabilisés

Un espace probabilisé est un cas particulier d'un espace mesuré pour lequel la masse totale de la mesure est égale à 1. L'objectif de cette section est de transcrire les notions introduites dans le Chapitre A en termes probabilistes. Notons que le point de vue est différent : on cherche à fournir un cadre mathématique pour décrire une expérience aléatoire. Ainsi

- l'ensemble sur lequel on travaille représente les résultats possibles dans l'expérience considérée ; on le note Ω .
- la σ -algèbre est donnée par les propriétés liées à l'expérience qui peuvent être vérifiées ; ce sont les parties de Ω dont on pourra évaluer la probabilité et on note cette collection \mathcal{F} .
- pour tout $A \in \mathcal{F}$, $\mathbb{P}(A)$ est une mesure de la confiance que l'on a en le fait que A va se réaliser ou non.

Définition B.1.1 Un *espace probabilisé* est un espace mesuré $(\Omega, \mathcal{F}, \mathbb{P})$ où la mesure \mathbb{P} est une mesure de probabilité, c'est-à-dire $\mathbb{P}(\Omega) = 1$.

Remarque B.1.2 Lorsque Ω est dénombrable, on travaillera en général avec $\mathcal{F} = \mathcal{P}(\Omega)$. Lorsque $\Omega = \mathbb{R}^d$, on prendra $\mathcal{F} = \mathcal{B}^d$.

Exemple B.1.3 (Lancer du dé) On considère l'ensemble $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de $\mathcal{P}(\Omega)$ et de la mesure $\mathbb{P} = \frac{1}{6} \sum_{n=1}^6 \delta_n$. Alors, l'espace $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est un espace probabilisé qui sert à modéliser le jet d'un dé. Comme attendu intuitivement, on a $\mathbb{P}(\{k\}) = \frac{1}{6}$ pour tout $k \in \Omega$ et la probabilité d'obtenir un nombre impair est $\mathbb{P}(\{1, 3, 5\}) = \frac{3}{6} = \frac{1}{2}$.

Exemple B.1.4 (Mesure discrète) Soit $\Omega = \{x_1, \dots, x_N\} \subseteq \mathbb{R}$ avec $N \in \mathbb{N} \cup \{\infty\}$ muni de $\mathcal{P}(\Omega)$. On considère la mesure de comptage

$$\mu : A \in \mathcal{P}(\Omega) \mapsto \# \{n \in \{1, \dots, N\} : x_n \in A\}.$$

On remarque que

$$\mu = \sum_{n=1}^N \delta_{x_n}.$$

Ce n'est pas une mesure de probabilité mais si N est fini, la mesure de probabilité associée est donnée par

$$\mu = \frac{1}{N} \sum_{n=1}^N \delta_{x_n}.$$

On retrouve le contexte *équiprobable* de l'exemple précédent. On peut généraliser l'exemple précédent en changeant les "poids" associés aux mesures de Dirac. Si $\alpha_1, \dots, \alpha_N$ sont des réels positifs tels que $\sum_{n=1}^N \alpha_n = 1$, alors la mesure

$$\sum_{n=1}^N \alpha_n \delta_{x_n}$$

est une mesure de probabilité.

Exemple B.1.5 (Expérience de Bernoulli) Il s'agit d'un cas particulier de l'exemple précédent. Soit $p \in [0, 1]$. On considère une expérience n'ayant que deux issues possibles : le succès avec une probabilité p et l'échec avec une probabilité $1 - p$. On code le succès par 1 et l'échec par 0. Ainsi, on a $\Omega = \{0, 1\}$ et sur $(\Omega, \mathcal{P}(\Omega))$, on considère la mesure $\mathbb{P} = p\delta_1 + (1 - p)\delta_0$. Lorsque $p = \frac{1}{2}$, cette mesure modélise le lancer d'une pièce dans un jeu de pile (codé par 0) ou face (codé par 1) équilibré.

Exemple B.1.6 (Mesure de Poisson) Si $\Omega = \mathbb{N}$ et $\mathcal{F} = \mathcal{P}(\Omega)$, on considère la mesure

$$\sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n$$

pour un paramètre $\lambda > 0$. Il s'agit d'une mesure de probabilité.

Exemple B.1.7 (Mesure à densité par rapport à λ) Si $f : \mathbb{R} \rightarrow [0, +\infty]$ est une application mesurable, alors l'application

$$\nu : A \in \mathcal{A} \mapsto \int_A f d\lambda = \int_A f(x) dx$$

est une mesure sur $(\mathbb{R}, \mathcal{B})$. On dit que f est une *densité* de ν par rapport à λ . Si f est de plus intégrable par rapport à λ et d'intégrale égale à 1, alors ν sera une mesure de probabilité.

Exemple B.1.8 (Mesure gaussienne) On considère la fonction f définie sur \mathbb{R} par

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Alors $\nu = f \cdot \lambda$ est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$. On a

$$\nu(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

pour tout ensemble borélien A de \mathbb{R} .

Exemple B.1.9 (Tirage continu uniforme) On considère l'expérience qui consiste à choisir au hasard un nombre réel entre 0 et 1. Par conséquent, $\Omega = [0, 1]$ et puisque la probabilité de tirer un nombre dans chaque intervalle est donnée par la longueur de cet intervalle, on considère naturellement l'espace probabilisé $([0, 1], \mathcal{B}([0, 1]), \lambda|_{[0, 1]})$.

Puisque les mesures de probabilité sont des mesures, elles vérifient les propriétés de la Proposition A.3.8. En particulier, puisque $\mathbb{P}(\Omega) = 1$, on a

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

pour tout $A \in \mathcal{F}$.

Définition B.1.10 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Un ensemble $A \in \mathcal{F}$ est appelé un *événement*. Un événement A a lieu \mathbb{P} -presque sûrement (\mathbb{P} -ps) s'il a lieu \mathbb{P} -presque partout, i.e. si $\mathbb{P}(A) = 1$.

En général et lorsque le contexte est clair, on dira que l'événement A (ou la propriété liée) a lieu presque sûrement, sans préciser la mesure de probabilité \mathbb{P} . La propriété suivante montre qu'une conjonction dénombrable de propriétés qui ont lieu presque sûrement est encore une propriété qui a lieu presque sûrement.

Proposition B.1.11 Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements qui ont lieu \mathbb{P} -presque sûrement. Alors $\bigcap_{n \in \mathbb{N}} A_n$ est un événement qui a lieu \mathbb{P} -presque sûrement.

Démonstration : Par hypothèse, on sait que $\mathbb{P}(A_n) = 1$ pour tout $n \in \mathbb{N}$ et par conséquent, $\mathbb{P}(A_n^c) = 0$ pour tout $n \in \mathbb{N}$. La sous- σ -additivité de \mathbb{P} implique alors que

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n^c\right) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n^c) = 0.$$

Puisque $(\bigcap_{n \in \mathbb{N}} A_n)^c = \bigcup_{n \in \mathbb{N}} A_n^c$, on obtient la conclusion. ■

B.2 Mesure de probabilité conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé lié à une expérience. Supposons que l'on sait qu'un événement B s'est produit. Si A est un autre événement, on cherche à calculer la probabilité de A avec l'information supplémentaire dont on dispose. On la note $\mathbb{P}(A|B)$.

Exemple B.2.1 Lors du lancer d'un dé, on sait que le résultat est un nombre pair, autrement dit $B = \{2, 4, 6\}$. Cette nouvelle information augmente notre confiance en le fait que l'événement $A = \{2\}$ se produise, et on estime intuitivement à présent sa probabilité à $1/3$ à la place de $1/6$. Donc

$$\mathbb{P}(A) = \frac{1}{6} \quad \text{et} \quad \mathbb{P}(A|B) = \frac{1}{3}.$$

Avant de définir proprement cette probabilité conditionnelle, regardons quelques cas.

- Si $A \cap B = \emptyset$ (événements *incompatibles*), alors il est clair que A ne peut pas se réaliser en même temps que B . Donc $\mathbb{P}(A|B) = 0$.

Exemple du dé : $B = \{2, 4, 6\}$ et $A = \{1, 3\}$.

— Si $A \cap B = B$, c'est-à-dire si $B \subseteq A$, alors il est *certain* que A va se réaliser. Donc $\mathbb{P}(A|B) = 1$.

Exemple du dé : $B = \{2, 4, 6\}$ et $A = \{2, 4, 5, 6\}$.

— Si $A \cap B \neq \emptyset$, alors l'événement A est réalisable, mais seulement les résultats de A qui sont dans B .

Exemple du dé : $B = \{2, 4, 6\}$ et $A = \{2, 3\}$.

Tout se passe comme si on avait *restreint* l'univers Ω à B . Etant donné un événement A , on ne s'intéresse plus qu'à $A \cap B$. On va donc regarder $\mathbb{P}(A \cap B)$. De plus, puisqu'on sait que B est réalisé, sa probabilité conditionnelle doit être égale à 1. On va donc normaliser $\mathbb{P}(A \cap B)$ par $\mathbb{P}(B)$.

Définition B.2.2 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{F}$ tel que $\mathbb{P}(B) \neq 0$. La *probabilité conditionnelle de A sachant B* est définie par

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Remarque B.2.3 Dans le cas équiprobable, la probabilité conditionnelle s'écrit

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{\#(A \cap B)}{\#\Omega}}{\frac{\#B}{\#\Omega}} = \frac{\#(A \cap B)}{\#B}.$$

Proposition B.2.4 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{F}$ tel que $\mathbb{P}(B) \neq 0$. L'application

$$\mathbb{P}(\cdot | B) : \mathcal{F} \rightarrow [0, 1]$$

est une mesure de probabilité.

Démonstration : Si $A \in \mathcal{F}$, on a

$$0 \leq \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

par monotonie de \mathbb{P} puisque $A \cap B \subseteq B$. Par conséquent, $\mathbb{P}(\cdot | B)$ est à valeurs dans $[0, 1]$. De plus, on a $\mathbb{P}(\Omega \cap B) = \mathbb{P}(B)$, d'où $\mathbb{P}(\Omega | B) = 1$. Enfin, si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{F} deux à deux disjoints, on a

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n | B\right) &= \frac{\mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right)}{\mathbb{P}(B)} \\ &= \frac{\sum_{n \in \mathbb{N}} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(A_n | B) \end{aligned}$$

puisque les ensembles $A_n \cap B$, $n \in \mathbb{N}$, sont deux à deux disjoints. ■

Les probabilités conditionnelles vérifient donc les propriétés classiques des mesures données à la Proposition A.3.8. De plus, en utilisant la définition de la probabilité conditionnelle, on obtient directement la règle de multiplication suivante. C'est cette règle de multiplication qui permet de justifier l'utilisation d'arbres pondérés pour calculer des probabilités.

Proposition B.2.5 (Règle de multiplication) *Pour tous $A_1, \dots, A_n \in \mathcal{F}$ tels que $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$, on a*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 \cap A_2) \cdots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Démonstration : On remarque que le membre de droite est égal à

$$\mathbb{P}(A_1) \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)} \frac{\mathbb{P}(A_3 \cap A_2 \cap A_1)}{\mathbb{P}(A_1 \cap A_2)} \cdots \frac{\mathbb{P}(A_n \cap A_{n-1} \cap \dots \cap A_1)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})}$$

qui se simplifie pour laisser $\mathbb{P}(A_n \cap A_{n-1} \cap \dots \cap A_1)$. ■

Exemple B.2.6 On tire 3 cartes sans remise d'un jeu de 52. Quelle est la probabilité d'avoir 3 as? Avec des notations évidentes :

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 \cap A_2) = \frac{4}{52} \frac{3}{51} \frac{2}{50}.$$

L'exemple ci-dessous illustre le fait que les probabilités conditionnelles nous donnent plus d'informations que les probabilités simples. Nous verrons qu'elles permettent de retrouver les probabilités non-conditionnelles via la formule des probabilités totales.

Exemple B.2.7 Supposons que parmi 200 étudiants, 60 sont des filles en math, 50 sont des filles en biologie, 60 sont des garçons en math et 30 sont des garçons en biologie.

	Math	Bio	
F	60	50	110
G	60	30	90
	120	80	200

On choisit un étudiant au hasard. Dans ce tableau, on lit que

$$\mathbb{P}(F \cap \text{Math}) = \frac{60}{200} = 0.3 \quad \text{et} \quad \mathbb{P}(\text{Math}) = \frac{120}{200} = 0.6.$$

La probabilité de choisir une fille sachant qu'on a choisi un mathématicien est donnée par la proportion de filles parmi les mathématiciens :

$$\mathbb{P}(F|\text{Math}) = \frac{60}{120} = 0.5.$$

On remarque que

$$\underbrace{\mathbb{P}(F|\text{Math})}_{=0.5} < \underbrace{\mathbb{P}(F)}_{=0.55} < \underbrace{\mathbb{P}(F|\text{Bio})}_{=0.625}.$$

La proportion de filles dans la population des biologistes est donc plus importante que la proportion de filles dans la population, et la proportion de filles dans la population des mathématiciens est moins importante.

On peut bien entendu réécrire la définition de la probabilité conditionnelle sous la forme

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A | B) = \mathbb{P}(A)\mathbb{P}(B | A)$$

d'où

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)}\mathbb{P}(B | A).$$

La définition de la probabilité conditionnelle est les formules ci-dessus mènent directement à deux résultats élémentaires mais très importants : la loi des probabilités totales et la formule de Bayes. Rappelons qu'une partition de Ω est une famille de sous-ensembles de Ω deux à deux disjoints et dont l'union donne Ω .

Proposition B.2.8 (Loi des probabilités totales) Soit $(B_n)_{n \in I}$ une partition finie ou dénombrable de Ω en événements tels que $\mathbb{P}(B_n) > 0$ pour tout $n \in I$. Alors pour tout $A \in \mathcal{F}$, on a

$$\mathbb{P}(A) = \sum_{n \in I} \mathbb{P}(A | B_n)\mathbb{P}(B_n).$$

Démonstration : Remarquons que A peut s'écrire comme l'union disjointe des événements $A \cap B_n$, $n \in I$. On obtient donc

$$\mathbb{P}(A) = \sum_{n \in I} \mathbb{P}(A \cap B_n) = \sum_{n \in I} \mathbb{P}(A | B_n)\mathbb{P}(B_n)$$

en utilisant la σ -additivité de \mathbb{P} . ■

Remarque B.2.9 On peut considérer que cette formule reste valide si $\mathbb{P}(B_n) = 0$ pour certains n car même si $\mathbb{P}(A | B_n)$ n'est pas bien défini, on le multiplie ensuite par 0.

Remarque B.2.10 En particulier, si $B \in \mathcal{F}$ est tel que $\mathbb{P}(B) \neq 0$ et $\mathbb{P}(B^c) \neq 0$, on trouve

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A | B)\mathbb{P}(B) + \mathbb{P}(A | B^c)\mathbb{P}(B^c)$$

qui est une moyenne pondérée de la probabilité de A sachant que B s'est produit et celle sachant que B ne s'est pas produit, les poids respectifs étant donnés par la probabilité que B se produise ou non.

Exemple B.2.11 Considérons un jeu de 52 cartes dont on tire 2 cartes sans remise. Quelle est la probabilité que la deuxième carte soit un as? On prend comme partition les deux événements $B_1 = \{\text{la première carte est un as}\}$ et $B_2 = \{\text{la première carte n'est pas un as}\}$. Notons maintenant l'événement d'intérêt $A = \{\text{la deuxième carte est un as}\}$. On déduit que

$$\mathbb{P}(A) = \mathbb{P}(A | B_1)\mathbb{P}(B_1) + \mathbb{P}(A | B_2)\mathbb{P}(B_2) = \frac{3}{51} \frac{4}{52} + \frac{4}{51} \frac{48}{52}.$$

Le résultat suivant est immédiat.

Corollaire B.2.12 (Règle de Bayes) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $A, B \in \mathcal{F}$. Alors

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} \mathbb{P}(A | B). \quad (\text{B.1})$$

Plus généralement, si $(B_n)_{n \in I}$ est une partition finie ou dénombrable de Ω en événements tels que $\mathbb{P}(B_n) > 0$ pour tout $n \in I$, alors

$$\mathbb{P}(B_k | A) = \frac{\mathbb{P}(B_k)}{\sum_{n \in I} \mathbb{P}(B_n)} \mathbb{P}(A | B_k). \quad (\text{B.2})$$

Exemple B.2.13 Supposons que la probabilité d'attraper une maladie D soit de 1%. Supposons également qu'il existe un test de dépistage pour la maladie avec les caractéristiques suivantes : le test rend un résultat positif 97% du temps si le patient est réellement atteint de la maladie D , et rend un résultat négatif 95% du temps si le patient n'est pas atteint de la maladie D . Si nous faisons le test sur une personne choisie au hasard :

- Quelle est la probabilité qu'elle soit indiquée comme malade ?
- Si le test rend un résultat positif, quelle est la probabilité que la personne soit réellement malade ?

On calcule

$$\mathbb{P}(T_+) = \mathbb{P}(D)\mathbb{P}(T_+ | D) + \mathbb{P}(D^c)\mathbb{P}(T_+ | D^c) \approx 0.06.$$

$$\mathbb{P}(D | T_+) = \frac{\mathbb{P}(D)}{\mathbb{P}(T_+)} \mathbb{P}(T_+ | D) \approx 0.16.$$

Remarquez notamment comment la fiabilité du test dépend de la prévalence de la maladie dans la population ! Si la maladie était plus rare il faudrait un test beaucoup plus efficace ! Avec une maladie rare (1 pour 10000) on a sous exactement les mêmes conditions : $\mathbb{P}(T_+) \approx 0,05$ et $\mathbb{P}(D | T_+) \approx 0,002$!

B.3 Indépendance d'événements

Un des concepts de base de la théorie des probabilités est celui de la dépendance (ou non-dépendance) d'événements entre eux. C'est avec la notion d'indépendance que la théorie des probabilités prend son autonomie par rapport à la théorie de la mesure. Philosophiquement, c'est un concept très complexe ; mathématiquement, c'est très simple : on dira que A et B sont dépendants si l'occurrence de B influence les chances de voir A , i.e. si $\mathbb{P}(A | B) \neq \mathbb{P}(A)$. Si ce n'est pas le cas, on dira A et B indépendants.

Définition B.3.1 Soient $A, B \in \mathcal{F}$. On dit que A et B sont *indépendants* si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Dans ce cas, on note $A \perp\!\!\!\perp B$.

La relation $\perp\!\!\!\perp$ est évidemment symétrique.

Remarque B.3.2 On pourrait également définir l'indépendance de A et B en demandant que $\mathbb{P}(A|B) = \mathbb{P}(A)$. Les avantages de la définition proposée sont qu'elle est valide même si $\mathbb{P}(B) = 0$ et qu'elle est symétrique en A et B .

Remarque B.3.3 Evidemment, pour tous $A, B \in \mathcal{F}$, si $\mathbb{P}(B) = 0$, alors $A \perp\!\!\!\perp B$. De plus, $A \perp\!\!\!\perp \Omega$.

Remarque B.3.4 Si $A, B \in \mathcal{F}$ sont incompatibles, i.e. si $A \cap B = \emptyset$, alors A et B ne peuvent pas être indépendants à moins que l'un des deux soit de probabilité nulle.

Exemple B.3.5 On tire une carte dans un jeu classique de 52 cartes. On considère les événements

- A = la carte tirée est un as,
- B = la carte tirée est un pique.

On remarque que $\mathbb{P}(A) = 4/52$, $\mathbb{P}(B) = 1/4$ et $\mathbb{P}(A \cap B) = 1/52 = \mathbb{P}(A)\mathbb{P}(B)$. Les événements A et B sont donc indépendants.

Remarque B.3.6 Attention, le fait que $A \perp\!\!\!\perp B$ et $B \perp\!\!\!\perp C$ n'implique pas que $A \perp\!\!\!\perp C$! Par exemple, reprenons l'exemple du jeu de cartes avec

- A = la carte tirée est un as,
- B = la carte tirée est un pique,
- C = la carte tirée est un roi.

On sait que A et B sont indépendants et par le même raisonnement, B et C sont indépendants. Mais bien sûr A et C ne le sont pas!

Remarque B.3.7

- On a vu que Ω est indépendant de Ω .
- On lance deux dés et on considère l'événement $A = \{\text{somme des points est } 7\}$ et l'événement $B = \{\text{premier résultat est } 6\}$. Alors

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = 1/6 \\ \mathbb{P}(B) &= \mathbb{P}(\{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}) = 1/6 \\ \mathbb{P}(A \cap B) &= \mathbb{P}(\{(6, 1)\}) = 1/36 = \mathbb{P}(A)\mathbb{P}(B)\end{aligned}$$

donc $A \perp\!\!\!\perp B$.

On étend la définition d'indépendance au cas de n événements.

Définition B.3.8 Soient $A_1, \dots, A_n \in \mathcal{F}$. On dit que des événements A_1, \dots, A_n sont *indépendants* si pour toute partie finie non-vide J de $\{1, \dots, n\}$, on a

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Cette définition est plus forte que de demander

$$\mathbb{P}\left(\bigcap_{j=1}^n A_j\right) = \prod_{j=1}^n \mathbb{P}(A_j).$$

C'est également plus fort que l'indépendance *deux-à-deux* pour laquelle on prend $\#J = 2$, c'est-à-dire on demande que

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \forall i \neq j.$$

Exemple B.3.9 Deux dés distinguables sont lancés. On considère les événements

- A = la somme des deux dés vaut 7,
- B = le résultat du premier dé est 4,
- C = le résultat du deuxième dé est 3.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

On a

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{6}{36} = \frac{1}{6}.$$

De plus,

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \frac{1}{36} = \mathbb{P}(B)\mathbb{P}(C)$$

mais

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{(4, 3)\}) = \frac{1}{36} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Annexe C

Théorème de Radon-Nikodym

Le Théorème de Radon-Nikodym est un outil indispensable de la théorie des probabilités.

C.1 Densités

Proposition C.1.1 Soient (X, \mathcal{A}, μ) un espace mesuré et $f : X \rightarrow [0, +\infty]$ une application mesurable. On définit l'application

$$\nu : \mathcal{A} \rightarrow [0, +\infty] : A \mapsto \int_A f d\mu.$$

Alors ν est une mesure sur (X, \mathcal{A}) et on dit que f est une densité de ν par rapport à μ . On note $\nu = f \cdot \mu$.

Démonstration : Bien sûr, $\nu(\emptyset) = 0$. De plus, si $(A_n)_{n \in \mathbb{N}}$ est une suite d'ensembles de \mathcal{A} deux à deux disjoints, alors

$$\nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \int_{\bigcup_{n \in \mathbb{N}} A_n} f d\mu = \int \sum_{n \in \mathbb{N}} f \mathbb{1}_{A_n} d\mu = \sum_{n \in \mathbb{N}} \int_{A_n} f d\mu = \sum_{n \in \mathbb{N}} \nu(A_n)$$

par le Corollaire A.5.13. ■

En théorie des probabilités, les lois de variables aléatoires \mathbb{P}_X sont souvent définies comme des densités par rapport à la mesure de Lebesgue. Par exemple, pour la loi normale, on a

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Proposition C.1.2 Soient (X, \mathcal{A}, μ) un espace mesuré, g une application mesurable de X dans $[0, +\infty]$ et f une application mesurable de X dans $\overline{\mathbb{R}}$. Alors f est $g \cdot \mu$ -intégrable si et seulement si fg est μ -intégrable et dans ce cas,

$$\int f d(g \cdot \mu) = \int fg d\mu.$$

Démonstration : Supposons tout d'abord que $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$ est une application de $\mathcal{S}^+(X, \mathcal{A})$. Alors

$$\int f d(g \cdot \mu) = \sum_{k=1}^n a_k g \cdot \mu(A_k) = \sum_{k=1}^n a_k \int_{A_k} g d\mu = \int fg d\mu$$

par linéarité de l'intégrale.

Si f est à valeurs dans $[0, +\infty]$, on peut considérer une suite croissante $(f_n)_{n \in \mathbb{N}}$ de $\mathcal{S}^+(X, \mathcal{A})$ qui converge vers f . Par la première partie de la preuve, il vient

$$\int f d(g \cdot \mu) = \lim_{n \rightarrow +\infty} \int f_n d(g \cdot \mu) = \lim_{n \rightarrow +\infty} \int f_n g d\mu$$

et le théorème de la convergence monotone permet de conclure.

Enfin, dans le cas général, il suffit de décomposer f en f^+ et f^- . ■

C.2 Mesures absolument continues et singulières

Soit (X, \mathcal{A}, μ) un espace mesuré et $f : X \rightarrow [0, +\infty]$ une application mesurable. On a vu dans la Proposition C.1.1 que l'application $\nu = f \cdot \mu$ définie par

$$\nu : \mathcal{A} \rightarrow [0, +\infty] : A \mapsto \int_A f d\mu$$

est une mesure sur (X, \mathcal{A}) . Remarquons que dans ce cas, si $A \in \mathcal{A}$ est tel que $\mu(A) = 0$, alors $g \mathbb{1}_A = 0$ μ -presque partout, d'où $\nu(A) = 0$.

Définition C.2.1 Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures sur (X, \mathcal{A}) . On dit que ν est *absolument continue* par rapport à μ si pour tout $A \in \mathcal{A}$ tel que $\mu(A) = 0$, on a $\nu(A) = 0$. Dans ce cas, on écrit

$$\nu \ll \mu.$$

Une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est *absolument continue* si elle est absolument continue par rapport à la mesure de Lebesgue.

Le but du Théorème de Radon-Nikodym est de démontrer que si ν est une mesure σ -finie absolument continue par rapport à μ , alors ν est de la forme $f \cdot \mu$. Il donne donc une réciproque au résultat présenté en début de section. Nous allons en outre montrer une décomposition de toute mesure σ -finie, appelée la décomposition de Lebesgue. Afin de présenter ce résultat, nous allons nous intéresser à la notion de singularité.

Définition C.2.2 Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures sur (X, \mathcal{A}) . On dit que μ et ν sont *mutuellement singulières* s'il existe un ensemble $A \in \mathcal{A}$ tel que

$$\mu(A) = 0 \quad \text{et} \quad \nu(A^c) = 0.$$

Dans ce cas, on écrit

$$\mu \perp \nu.$$

Une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est *singulière* si elle est mutuellement singulière avec la mesure

de Lebesgue.

Exemple C.2.3 Si μ est une mesure finie discrète sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, c'est-à-dire il existe D dénombrable tel que $\mu(D^c) = 0$, alors μ est singulière puisque $\mathcal{L}(D) = 0$.

Remarquons que dans ce cas, si $B \in \mathcal{A}$ est tel que $\nu(B) > 0$, alors $B \cap A \neq \emptyset$. Ainsi, ν concentre son effet sur l'ensemble A qui est de μ mesure nulle. À l'opposé, si $\nu \ll \mu$, alors ν n'a aucun effet sur les ensembles de μ mesure nulle. Il s'agit donc de deux notions qui s'excluent.

Proposition C.2.4 Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures sur (X, \mathcal{A}) . Si $\nu \ll \mu$ et $\nu \perp \mu$, alors $\nu = 0$.

Démonstration : Soit $A \in \mathcal{A}$ tel que $\mu(A) = 0$ et $\nu(A^c) = 0$. Si $B \in \mathcal{A}$ est un sous-ensemble de A , on a $\mu(B) \leq \mu(A) = 0$ et donc puisque $\nu \ll \mu$, $\nu(B) = 0$. Si $B \in \mathcal{A}$ est un sous-ensemble de A^c , on a $\nu(B) \leq \nu(A^c) = 0$ et donc $\nu(B) = 0$. Ainsi, pour tout $B \in \mathcal{A}$, on trouve

$$\nu(B) = \nu(A \cap B) + \nu(A^c \cap B) = 0.$$

■

La proposition suivante donne quelques relations supplémentaires entre les notions d'absolue continuité et de singularité.

Proposition C.2.5 Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν_1, ν_2 des mesures sur (X, \mathcal{A}) .

1. Si $\nu_1 \perp \mu$ et $\nu_2 \perp \mu$, alors $\nu_1 + \nu_2 \perp \mu$.
2. Si $\nu_1 \ll \mu$ et $\nu_2 \perp \mu$, alors $\nu_1 \perp \nu_2$.

Démonstration : 1. Soient $A_1, A_2 \in \mathcal{A}$ tels que $\mu(A_1) = 0$, $\nu_1(A_1^c) = 0$ et $\mu(A_2) = 0$, $\nu_2(A_2^c) = 0$. Alors $A_1 \cup A_2$ est tel que

$$\mu(A_1 \cup A_2) = 0$$

et

$$(\nu_1 + \nu_2)((A_1 \cup A_2)^c) = \nu_1(A_1^c \cap A_2^c) + \nu_2(A_1^c \cap A_2^c) \leq \nu_1(A_1^c) + \nu_2(A_2^c) = 0.$$

2. Soit $A \in \mathcal{A}$ tel que $\mu(A) = 0$ et $\nu_2(A^c) = 0$. Alors puisque $\nu_1 \ll \mu$, on a aussi $\nu_1(A) = 0$, ce qui suffit. ■

C.3 Décomposition de Lebesgue et Théorème de Radon-Nikodym

L'objectif de cette section est de démontrer que toute mesure σ -finie se décompose en la somme d'une mesure absolument continue et d'une mesure singulière par rapport à une autre mesure σ -finie donnée.

Dans un premier temps, on montre un résultat pour deux mesures finies et ordonnées.

Lemme C.3.1 Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures finies sur (X, \mathcal{A}) telles que $\nu \leq \mu$. Il existe une application mesurable $f : X \rightarrow [0, 1]$ telle que pour tout $A \in \mathcal{A}$

$$\nu(A) = \int_A f d\mu.$$

Démonstration : Considérons l'ensemble \mathcal{H} des applications mesurables $h : X \rightarrow [0, +\infty]$ telles que

$$\int_A h d\mu \leq \nu(A)$$

pour tout $A \in \mathcal{A}$. Montrons qu'il existe une application $f \in \mathcal{H}$ à valeurs dans $[0, 1]$ telle que

$$\int f d\mu = \sup \left\{ \int h d\mu : h \in \mathcal{H} \right\}.$$

Nous montrerons ensuite que cette application satisfait la thèse.

Bien sûr, $\mathcal{H} \neq \emptyset$ puisqu'il contient la fonction nulle. De plus, \mathcal{H} est stable par passage au maximum : en effet, soient $h_1, h_2 \in \mathcal{H}$ et $A \in \mathcal{A}$. Si $A_1 = \{x \in A : h_1(x) > h_2(x)\}$, alors

$$\int_A \max\{h_1, h_2\} d\mu = \int_{A_1} h_1 d\mu + \int_{A \setminus A_1} h_2 d\mu \leq \nu(A_1) + \nu(A \setminus A_1) = \nu(A).$$

Ainsi, $\max\{h_1, h_2\} \in \mathcal{H}$.

Soit $(h_n)_{n \in \mathbb{N}}$ une suite de \mathcal{H} telle que

$$\lim_{n \rightarrow +\infty} \int h_n d\mu = \sup \left\{ \int h d\mu : h \in \mathcal{H} \right\}.$$

Quitte à remplacer h_n par $\max\{h_1, \dots, h_n\} \in \mathcal{H}$, on peut supposer que la suite $(h_n)_{n \in \mathbb{N}}$ est croissante. En particulier, on peut considérer $f = \lim_{n \rightarrow +\infty} h_n$. Le théorème de la convergence monotone donne

$$\int_A f d\mu = \lim_{n \rightarrow +\infty} \int_A h_n d\mu \leq \nu(A)$$

pour tout $A \in \mathcal{A}$. Par conséquent, $f \in \mathcal{H}$. En prenant $A = X$, on a également

$$\int f d\mu = \sup \left\{ \int h d\mu : h \in \mathcal{H} \right\}.$$

Remarquons de plus que $f \leq 1$ μ -presque partout. En effet, par le Lemme A.5.10, on sait que pour tout $n \in \mathbb{N}_0$,

$$(1 + 1/n)\mu(A_n) \leq \int_{A_n} f d\mu \leq \nu(A_n) \leq \mu(A_n)$$

si $A_n = \{x \in X : f(x) \geq 1 + 1/n\}$. Par conséquent, $\mu(A_n) = 0$ et donc

$$\mu(\{x \in X : f(x) > 1\}) \leq \sum_{n \in \mathbb{N}_0} \mu(A_n) = 0.$$

Ainsi, quitte à remplacer f par une application égale à f μ -presque partout, on peut supposer que $f \leq 1$.

Remarquons à présent que si $h : X \rightarrow [0, +\infty]$ est une application mesurable pour laquelle il existe $A \in \mathcal{A}$ tel que

$$\int_B h d\mu \leq \nu(B)$$

pour tout $B \in \mathcal{A}$, $B \subseteq A$, alors $h \leq f$ μ -presque partout sur A . En effet, considérons l'ensemble $B = \{x \in A : h(x) > f(x)\}$. Alors, on a $h\mathbb{1}_A \in \mathcal{H}$ et donc $\max\{h\mathbb{1}_A, f\} \in \mathcal{H}$. Il s'ensuit que

$$\int \max\{h\mathbb{1}_A, f\} d\mu \leq \int f d\mu = \int_B f d\mu + \int_{X \setminus B} f d\mu.$$

D'autre part, on a également

$$\int \max\{h\mathbb{1}_A, f\} d\mu = \int_B h d\mu + \int_{X \setminus B} f d\mu.$$

Par conséquent, on obtient

$$\int_B f d\mu \geq \int_B h d\mu$$

et on en tire que

$$\int_B f - h d\mu \geq 0.$$

Comme on a $f - h < 0$ sur B , cela n'est possible que si $\mu(B) = 0$, d'où $h \leq f$ μ -presque partout sur A .

Montrons à présent que f convient, c'est-à-dire que pour tout $A \in \mathcal{A}$, on a

$$\nu(A) = \int_A f d\mu.$$

Par définition, on sait déjà que

$$\nu(A) \geq \int_A f d\mu$$

et il suffit de montrer l'autre inégalité. Bien sûr, puisque $\nu \leq \mu$, il suffit de le montrer pour tout $A \in \mathcal{A}$ tel que $\mu(A) > 0$. Soit $\varepsilon > 0$. Posons

$$P_\varepsilon = \left\{ A \in \mathcal{A} : \int_A (f + \varepsilon) d\mu > \nu(A) \right\}.$$

Remarquons que si $A \in \mathcal{A}$ est tel que $\mu(A) > 0$, alors il existe $B \subseteq A$ tel que $B \in P_\varepsilon$ (et donc $\mu(B) > 0$). En effet, sinon, on aurait

$$\int_B (f + \varepsilon) d\mu \leq \nu(B)$$

pour tout $B \subseteq A$, et donc par ce qui précède, on en tirerait que $f + \varepsilon \leq f$ μ -presque partout sur A . Cela n'est possible que si $\mu(A) = 0$, d'où une contradiction.

Fixons $A \in \mathcal{A}$ tel que $\mu(A) > 0$ et posons

$$\alpha_0 = \sup\{\mu(B) : B \in P_\varepsilon, B \subseteq A\}.$$

On peut trouver $n_0 \in \mathbb{N}_0$ et $B_0 \in P_\varepsilon$, $B_0 \subseteq A$, tels que

$$\frac{1}{n_0} < \alpha_0 \leq \frac{1}{n_0 - 1} \quad \text{et} \quad \frac{1}{n_0} < \mu(B_0)$$

(si $n_0 = 1$, on utilise la convention que $\frac{1}{n_0 - 1} = +\infty$). Soit à présent $A_1 = A \setminus B_0$. Si $\mu(A_1) = 0$, on s'arrête. Sinon, soit

$$\alpha_1 = \sup\{\mu(B) : B \in P_\varepsilon, B \subseteq A_1\},$$

et $n_1 \in \mathbb{N}_0$, $B_1 \in P_\varepsilon$, $B_1 \subseteq A_1$, tels que

$$\frac{1}{n_1} < \alpha_1 \leq \frac{1}{n_1 - 1} \quad \text{et} \quad \frac{1}{n_1} < \mu(B_1).$$

On pose alors $A_2 = A_1 \setminus B_1$. Par induction, on construit ainsi des suites $(A_i)_{i \in \{1, \dots, J\}}$, $(B_i)_{i \in \{0, \dots, J\}}$, $(n_i)_{i \in \{0, \dots, J\}}$ et $(\alpha_i)_{i \in \{0, \dots, J\}}$ avec $J \in \mathbb{N} \cup \{+\infty\}$ (le procédé peut ne jamais s'arrêter). Posons

$$A_\infty = A \setminus \bigcup_{i \in \{0, \dots, J\}} B_i$$

et montrons que cet ensemble est μ -négligeable.

Si la récurrence s'est terminée en un temps fini, on a $A_\infty = A_k$ pour le premier k tel que $\mu(A_k) = 0$.

Sinon, puisque les B_i sont des sous-ensembles de A deux à deux disjoints, on a

$$\sum_{i \in \mathbb{N}} \frac{1}{n_i} \leq \sum_{i \in \mathbb{N}} \mu(B_i) \leq \mu(A) < +\infty$$

et donc $\lim_{i \rightarrow +\infty} n_i = +\infty$. Par conséquent, on a également $\lim_{i \rightarrow +\infty} \alpha_i = 0$. Soient $B \subseteq A_\infty$ et $B \in P_\varepsilon$. Puisque pour tout entier $i \in \mathbb{N}$, on a aussi $B \subset A_i$, on en tire que donc $\mu(B) \leq \alpha_i$. Par passage à la limite, on conclut que $\mu(B) = 0$. Par conséquent, les sous-ensembles de A_∞ qui appartiennent à P_ε sont μ -négligeables. Cela n'est possible que si $\mu(A_\infty) = 0$.

Dans les deux cas, puisque $A = \bigcup_i B_i \cup A_\infty$ et que les ensembles sont deux à deux disjoints, on en déduit que

$$\int_A (f + \varepsilon) d\mu = \sum_{i \in \mathbb{N}} \int_{B_i} (f + \varepsilon) d\mu > \sum_{k \in \mathbb{N}} \nu(B_k) = \nu(A).$$

Cela signifie que $A \in P_\varepsilon$.

En conclusion, pour tout $A \in \mathcal{A}$ tel que $\mu(A) > 0$ et pour tout $\varepsilon > 0$, on a

$$\int_A (f + \varepsilon) d\mu > \nu(A)$$

et comme $\varepsilon > 0$ est arbitraire, il vient

$$\int_A f d\mu \geq \nu(A),$$

ce qui conclut la preuve. ■

Théorème C.3.2 Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures σ -finies sur (X, \mathcal{A}) . Il existe un ensemble $D \in \mathcal{A}$ et une application mesurable $f : X \rightarrow [0, +\infty]$ tels que

$$\mu(D) = 0 \quad \text{et} \quad \forall A \in \mathcal{A}, \quad \nu(A) = \nu(A \cap D) + \int_A f d\mu.$$

Démonstration : Démontrons tout d'abord le théorème dans le cas où μ et ν sont deux mesures finies. Considérons la mesure $\lambda = \nu + \mu$. Alors $\nu \leq \lambda$ et le Lemme C.3.1 nous donne une application mesurable $g : X \rightarrow [0, 1]$ telle que

$$\nu(A) = \int_A g d\lambda = \int_A g d\nu + \int_A g d\mu$$

pour tout $A \in \mathcal{A}$.

D'une part, puisque $\nu(A) = \int_A 1 d\nu$, on en tire que

$$\int_A (1 - g) d\nu = \int_A g d\mu$$

pour tout $A \in \mathcal{A}$. Par linéarité, on en tire que pour toute application $h \in \mathcal{S}^+(X, \mathcal{A})$, on a

$$\int (1 - g)h d\nu = \int gh d\mu.$$

En utilisant le théorème de la convergence monotone, l'égalité précédente est vérifiée également pour toute fonction mesurable positive h .

D'autre part, on a aussi

$$\mu(A) = \lambda(A) - \nu(A) = \int_A 1 d\lambda - \int_A g d\lambda = \int_A (1 - g) d\lambda.$$

Posons $D = \{x \in X : g(x) = 1\}$. Alors $D \in \mathcal{A}$ et l'égalité précédente implique que $\mu(D) = 0$. Notons $C = X \setminus D$ et posons

$$h = \mathbb{1}_C \frac{1}{1 - g}.$$

C'est une application mesurable positive. Si $A \in \mathcal{A}$, on a alors

$$\nu(A \cap C) = \int_{A \cap C} (1 - g)h d\nu = \int_{A \cap C} gh d\mu.$$

Pour conclure, il suffit donc de prendre $f = gh = \mathbb{1}_C \frac{g}{1 - g}$.

Considérons à présent le cas de deux mesures σ -finies. On considère une suite $(A_n)_{n \in \mathbb{N}}$ d'ensembles deux à deux disjoints de \mathcal{A} telle que

$$X = \bigcup_{n \in \mathbb{N}} A_n, \quad \mu(A_n) < +\infty \quad \text{et} \quad \nu(A_n) < +\infty$$

(Remarquons que ceci est toujours possible : en effet, si $X = \bigcup_{n \in \mathbb{N}} X_n$ et $X = \bigcup_{n \in \mathbb{N}} Y_n$ avec $\mu(X_n) < +\infty$, $\nu(Y_n) < +\infty$, alors $X = \bigcup_{m,n} X_n \cap Y_m$). On considère ensuite, pour tout $n \in \mathbb{N}$, les mesures μ_n et ν_n définies par

$$\mu_n(A) = \mu(A \cap A_n) \quad \text{et} \quad \nu_n(A) = \nu(A \cap A_n)$$

pour tout $A \in \mathcal{A}$. Comme les mesures μ_n et ν_n sont finies, la première partie de la preuve nous donne un ensemble $D_n \in \mathcal{A}$ et une application mesurable positive f_n tels que

$$\mu_n(D_n) = 0 \quad \text{et} \quad \forall A \in \mathcal{A}, \quad \nu_n(A) = \nu_n(A \cap D_n) + \int_A f_n d\mu_n.$$

Posons $f = \sum_{n \in \mathbb{N}} f_n \mathbb{1}_{A_n}$. Alors f est une application mesurable positive et comme les A_n sont deux à deux disjoints, on a

$$\begin{aligned} \nu(A) &= \nu\left(\bigcup_{n \in \mathbb{N}} A \cap A_n\right) = \sum_{n \in \mathbb{N}} \nu(A \cap A_n) \\ &= \sum_{n \in \mathbb{N}} \left(\nu_n(A \cap D_n) + \int_A f_n d\mu_n\right) \\ &= \sum_{n \in \mathbb{N}} \nu(A \cap D_n \cap A_n) + \sum_{n \in \mathbb{N}} \int_A f_n \mathbb{1}_{A_n} d\mu \\ &= \nu\left(\bigcup_{n \in \mathbb{N}} A \cap D_n \cap A_n\right) + \int_A \sum_{n \in \mathbb{N}} f_n \mathbb{1}_{A_n} d\mu \\ &= \nu\left(A \cap \left(\bigcup_{n \in \mathbb{N}} D_n \cap A_n\right)\right) + \int_A f d\mu \end{aligned}$$

en utilisant le Corollaire A.5.13. Pour conclure, il suffit de poser $D = \bigcup_{n \in \mathbb{N}} D_n \cap A_n$ et de remarquer que

$$\mu(D) = \sum_{n \in \mathbb{N}} \mu(D_n \cap A_n) = \sum_{n \in \mathbb{N}} \mu_n(D_n) = 0.$$

■

Présentons à présent les deux corollaires importants du théorème précédent.

Théorème C.3.3 (Théorème de Radon-Nikodym) *Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures σ -finies sur (X, \mathcal{A}) . Alors $\nu \ll \mu$ si et seulement si il existe une application mesurable $f : X \rightarrow [0, +\infty]$ telle que*

$$\nu(A) = \int_A f d\mu$$

pour tout $A \in \mathcal{A}$, c'est-à-dire $\nu = f \cdot \mu$. De plus, si g est une autre application telle que $\nu = g \cdot \mu$, alors $f = g$ μ -presque partout.

Démonstration : On sait que si $\nu = f \cdot \mu$, alors $\nu \ll \mu$. Réciproquement, supposons que $\nu \ll \mu$. Par le Théorème C.3.2, il existe $D \in \mathcal{A}$ et $f : X \rightarrow [0, +\infty]$ mesurable tels que

$$\mu(D) = 0 \quad \text{et} \quad \forall A \in \mathcal{A}, \quad \nu(A) = \nu(A \cap D) + \int_A f d\mu.$$

Alors on a aussi $\nu(D) = 0$, et donc

$$\nu(A) = \int_A f d\mu.$$

pour tout $A \in \mathcal{A}$.

Soit g une autre application mesurable telle que $\nu = g \cdot \mu$. Supposons dans un premier temps que ν est finie. Alors les applications f et g sont μ -intégrables et donc finies μ -presque partout. Soit $N \in \mathcal{A}$ un ensemble μ -négligeable en dehors duquel f et g sont finies. Considérons l'application mesurable h définie par

$$h(x) = (f(x) - g(x)) \mathbf{1}_{N^c}.$$

On a

$$\int_A h d\mu = 0$$

pour tout $A \in \mathcal{A}$. Soit $A = \{x \in X : h(x) > 0\}$. La relation précédente nous donne que $\mu(A) = 0$. De même, $\mu(\{x \in X : h(x) < 0\}) = 0$. Par conséquent, $h = 0$ μ -presque partout, d'où $f = g$ μ -presque partout.

Si ν est σ -fini, soit $(A_n)_{n \in \mathbb{N}}$ une suite d'ensembles deux à deux disjoints de \mathcal{A} telle que

$$X = \bigcup_{n \in \mathbb{N}} A_n \quad \text{et} \quad \nu(A_n) < +\infty.$$

L'argument précédent montre que $f = g$ μ -presque partout sur chaque ensemble A_n . Puisqu'une union dénombrable d'ensembles μ -négligeables est μ -négligeable, on obtient la conclusion. ■

Définition C.3.4 La fonction f intervenant dans le théorème précédent est appelée une *densité* ou une *dérivée de Radon-Nikodym* de ν par rapport à μ . Comme elle est unique à un ensemble μ -négligeable près, on parle aussi de la dérivée de Radon-Nikodym de ν par rapport à μ et on la note

$$f = \frac{d\nu}{d\mu}.$$

Ainsi, avec les notations précédemment introduites, on a

$$\nu = \frac{d\nu}{d\mu} \cdot \mu.$$

Théorème C.3.5 (Décomposition de Lebesgue) Soit (X, \mathcal{A}) un espace mesurable et soient μ, ν deux mesures σ -finies sur (X, \mathcal{A}) . Il existe un unique couple (ν_a, ν_s) de mesures sur (X, \mathcal{A}) telles que

$$\nu_a \ll \mu, \quad \nu_s \perp \mu \quad \text{et} \quad \nu = \nu_a + \nu_s.$$

On appelle le couple (ν_a, ν_s) la décomposition de Lebesgue de ν par rapport à μ . On dit que ν_a est la partie absolument continue de ν par rapport à μ , et que ν_s est la partie singulière de ν par rapport à μ .

Démonstration : Par le Théorème C.3.2, il existe $D \in \mathcal{A}$ et $f : X \rightarrow [0, +\infty]$ mesurable tels que

$$\mu(D) = 0 \quad \text{et} \quad \forall A \in \mathcal{A}, \quad \nu(A) = \nu(A \cap D) + \int_A f d\mu.$$

Posons

$$\nu_s(A) = \nu(A \cap D) \quad \text{et} \quad \nu_a(A) = \int_A f d\mu.$$

Bien sûr on a $\nu = \nu_s + \nu_a$. Par définition, on a aussi $\nu_s(D^c) = 0$ et donc $\nu_s \perp \mu$. Enfin, on sait que $\nu_a \ll \mu$.

Il reste à montrer l'unicité de la décomposition. Supposons que les couples (ν_a, ν_s) et (ν'_a, ν'_s) satisfont l'hypothèse. Alors $\nu_a + \nu_s = \nu'_a + \nu'_s$, d'où

$$\nu_a - \nu'_a = \nu'_s - \nu_s.$$

Posons $\lambda = \nu_a - \nu'_a = \nu'_s - \nu_s$. On procède comme dans les Propositions C.2.4 et C.2.5 pour montrer qu'alors $\lambda = 0$ (on ne s'y ramène pas directement car λ n'est pas nécessairement positive). Soient $D, D' \in \mathcal{A}$ tels que $\mu(D) = \mu(D') = 0$ et $\nu_s(D^c) = \nu'_s(D'^c) = 0$. Posons $B = D \cup D'$. Ainsi, $\mu(B) = 0$ et donc pour tout $A \in \mathcal{A}$, on a $\nu_a(A \cap B) = 0$ puisque $\nu_a \ll \mu$. Par conséquent,

$$\nu_a(A) = \nu_a(A \cap B) + \nu_a(A \cap B^c) = \nu_a(A \cap B^c).$$

On procède de même pour ν'_a . Il vient alors

$$\begin{aligned} \lambda(A) = \nu_a(A) - \nu'_a(A) &= \nu_a(A \cap B^c) - \nu'_a(A \cap B^c) \\ &= \nu_s(A \cap B^c) - \nu'_s(A \cap B^c) \\ &= 0 \end{aligned}$$

puisque $A \cap B^c = A \cap D^c \cap D'^c$ qui est un sous-ensemble de D^c et de D'^c . ■

Terminons ce chapitre par un raffinement de la décomposition de Lebesgue d'une mesure finie définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition C.3.6 Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que $x \in \mathbb{R}$ est un *atome* de μ si $\mu(\{x\}) > 0$. La mesure μ est *purement atomique* s'il existe $S \in \mathcal{B}(\mathbb{R})$ tel que $\mu(S^c) = 0$ et tout $x \in S$ est un atome de μ .

Par exemple, toute mesure de Dirac est purement atomique.

Proposition C.3.7 Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors l'ensemble des atomes de μ est au plus dénombrable.

Démonstration : Soit $S = \{x \in \mathbb{R} : \mu(\{x\}) > 0\}$. Pour tout $n \in \mathbb{N}$, il existe un nombre fini de $x \in E$ tels que $\mu(\{x\}) \geq 1/n$ (sinon la mesure de cet ensemble serait infinie). Puisque

$$S = \bigcup_{n \in \mathbb{N}_0} \{x \in \mathbb{R} : \mu(\{x\}) \geq 1/n\},$$

on en tire que S est au plus dénombrable. ■

Corollaire C.3.8 Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors μ est purement atomique si et seulement si μ est de la forme

$$\mu = \sum_{n \in \mathbb{N}} \alpha_n \delta_{x_n}$$

où $(\alpha_n)_{n \in \mathbb{N}}$ est une suite de réels positifs ou nuls, et $(x_n)_{n \in \mathbb{N}}$ est une suite de réels.

Démonstration : Supposons que μ est purement atomique. Alors l'ensemble S des atomes de μ est au plus dénombrable et $\mu(S^c) = 0$. Clairement,

$$\mu(A) = \mu(A \cap S) + \mu(A \cap S^c) = \sum_{x \in S} \mu(A \cap \{x\}) = \sum_{x \in X} \mu(\{x\}) \delta_x(A).$$

La réciproque est triviale. ■

Ainsi, les mesures purement atomiques sont les mesures discrètes. A l'opposé, on a la définition suivante.

Définition C.3.9 Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que μ est *diffuse* si elle ne possède pas d'atome, c'est-à-dire $\mu(\{x\}) = 0$ pour tout $x \in \mathbb{R}$.

Par exemple, la mesure de Lebesgue est diffuse. Toute mesure absolument continue par rapport à la mesure de Lebesgue est également diffuse.

Remarque C.3.10 Les mesures diffuses sont parfois appelées *mesures continues* car on peut montrer que si μ est diffuse, alors pour tout $\alpha \in [0, \mu(\mathbb{R})]$, il existe $A \in \mathcal{B}(\mathbb{R})$ tel que $\mu(A) = \alpha$.

Bien sûr, toute mesure μ purement atomique et diffuse est identiquement nulle : en effet, si μ est purement atomique, il existe $S \in \mathcal{B}(\mathbb{R})$ tel que $\mu(S^c) = 0$ et tout $x \in S$ est un atome de μ . Mais comme μ est diffuse, $S = \emptyset$ et donc $S^c = \mathbb{R}$.

Remarquons également que si μ est une mesure diffuse, alors $\mu \perp \delta_x$ pour tout $x \in \mathbb{R}$. En effet, $\mu(\{x\}) = 0$ et $\delta_x(\mathbb{R} \setminus \{x\}) = 0$. En fait, en utilisant le Corollaire C.3.8, on a que toute mesure diffuse est singulière par rapport à une mesure purement atomique.

Proposition C.3.11 Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors, il existe un triplet (μ_d, μ_s, μ_a) de mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles que μ_d est une mesure discrète, μ_s est une mesure diffuse singulière par rapport à la mesure de Lebesgue, μ_a est une mesure absolument continue par rapport à la mesure de Lebesgue, et

$$\mu = \mu_d + \mu_s + \mu_a.$$

Démonstration : Soit S l'ensemble des atomes de μ . Comme S est au plus dénombrable, la mesure μ_d définie par

$$\mu_d(A) = \mu(A \cap S)$$

est discrète. Considérons à présent la décomposition de Lebesgue (μ_s, μ_a) de la mesure définie par $A \mapsto \mu(A \cap S^c)$ par rapport à la mesure de Lebesgue. Alors

$$\mu = \mu_d + \mu_s + \mu_a$$

et il reste à montrer que la mesure μ_s est diffuse. C'est évident puisque $\mu_s(\{x\}) \leq \mu(\{x\} \cap S^c) = 0$. ■

Cette dernière décomposition est beaucoup utilisée en probabilités. On distingue trois types disjoints de lois :

- les lois discrètes : ce sont les lois purement atomiques,

- les lois absolument continues : ce sont les lois qui admettent une densité par rapport à la mesure de Lebesgue,
- les lois continues singulières : ce sont les lois diffuses, donc qui ne possèdent pas d'atomes, qui sont singulières par rapport à la mesure de Lebesgue.

Toute loi absolument continue est diffuse, mais l'implication inverse est fautive : la loi de Cantor (dont la fonction de répartition est donnée par l'escalier du diable) est un exemple de loi continue qui est singulière par rapport à la mesure de Lebesgue.

De manière générale, la proposition précédente montre que toute loi de probabilité est une loi mixte donnée par la somme d'une loi discrète, d'une loi continue singulière et d'une loi absolument continue.

Bibliographie

- [1] Barbe, Ph. and Ledoux, M. (2021) *Probabilité*, EDP Sciences.
- [2] Billingsley, Patrick (1999) *Convergence of probability measures*, New York; John Wiley & Sons.
- [3] Casella, G. and Berger, R. L. (2002) *Statistical inference, Volume 2* Duxbury Pacific Grove.
- [4] Cohn, S. L. (1980) *Measure Theory*, Birkhäuser, Boston.
- [5] Durrett, R. (2005) *Probability : Theory and Examples*, 3rd edition, Duxbury.
- [6] Le Gall, J.F. (2006) *Intégration, probabilités et processus aléatoires*, Notes de cours, École Normale Supérieure de Paris.
- [7] Li, D., *Probabilités*, Notes de cours, Université d'Artois.
- [8] Matheron, E., *Probabilités*, Notes de cours, Université de Lens.
- [9] Steele, J. M. *Stochastic calculus and Financial Applications*, Vol 45. Stochastic Modeling and Applied Probability, Springer.
- [10] Swan, Y. (2021) *Un premier cours de probabilité*, Notes de cours, ULB.
- [11] Williams, D. (1991) *Probability with Martingales*, Cambridge University Press.